Virtual Observatories: How Modern Technology is Reshaping Research in Astronomy

Joseph C. Jacob
Parallel Applications Technologies (3677)
Jet Propulsion Laboratory
California Institute of Technology

JPL Section 367 All-Lab Lecture
May 13, 2003

Acknowledgements

Contributors

A. Bergou (3676), G. B. Berriman (IPAC), G. Block (3677), L. Brieger (SDSC), R. Brunner (UIUC), J. Collier (3677), L. Craymer (3660), D. Curkendall (3670), S. G. Djorgovski (CIT), J. Good (IPAC), L. E. Husman (3677), M. Judd (3200), D. S. Katz (3677), G. Kremenek (SDSC), A. C. Laity (IPAC), P. P. Li (3677), C. Lonsdale (IPAC), A. Mahabal (CIT), C. Miller (3677), R. Moore (SDSC), L. Plesea (3677), T. Prince (CIT, JPL), H. Siegel (3677), A. Szalay (JHU), R. Williams (CACR), National Virtual Observatory Project, Supercomputer and High Definition Systems Group (366), Enterprise Networks and Telecommunications Group (366), NASA Research and Education Network (NREN) Project

Sponsors

NASA Earth Science Technology Office (ESTO) – Computational Technologies (CT) Project

NASA Space Science Applications of Information Technology (SAIT) Program, Office of Space Science (OSS)

NASA Ames Research Center (ARC), Computing Networking and Information Systems (CNIS) Program

National Science Foundation (NSF)

Outline

- Virtual Observatories (VO)
  - Motivation
  - National Virtual Observatory
- yourSky Custom Mosaic Server
- Montage: High Science Quality Mosaics
- Visualization
  - High Performance / Powerwall
  - Web-Based / Desktop Access
- Computational Grids

The Data Avalanche!

Growth in Aperture & Focal Plane
Of Institutionally Managed Observatories

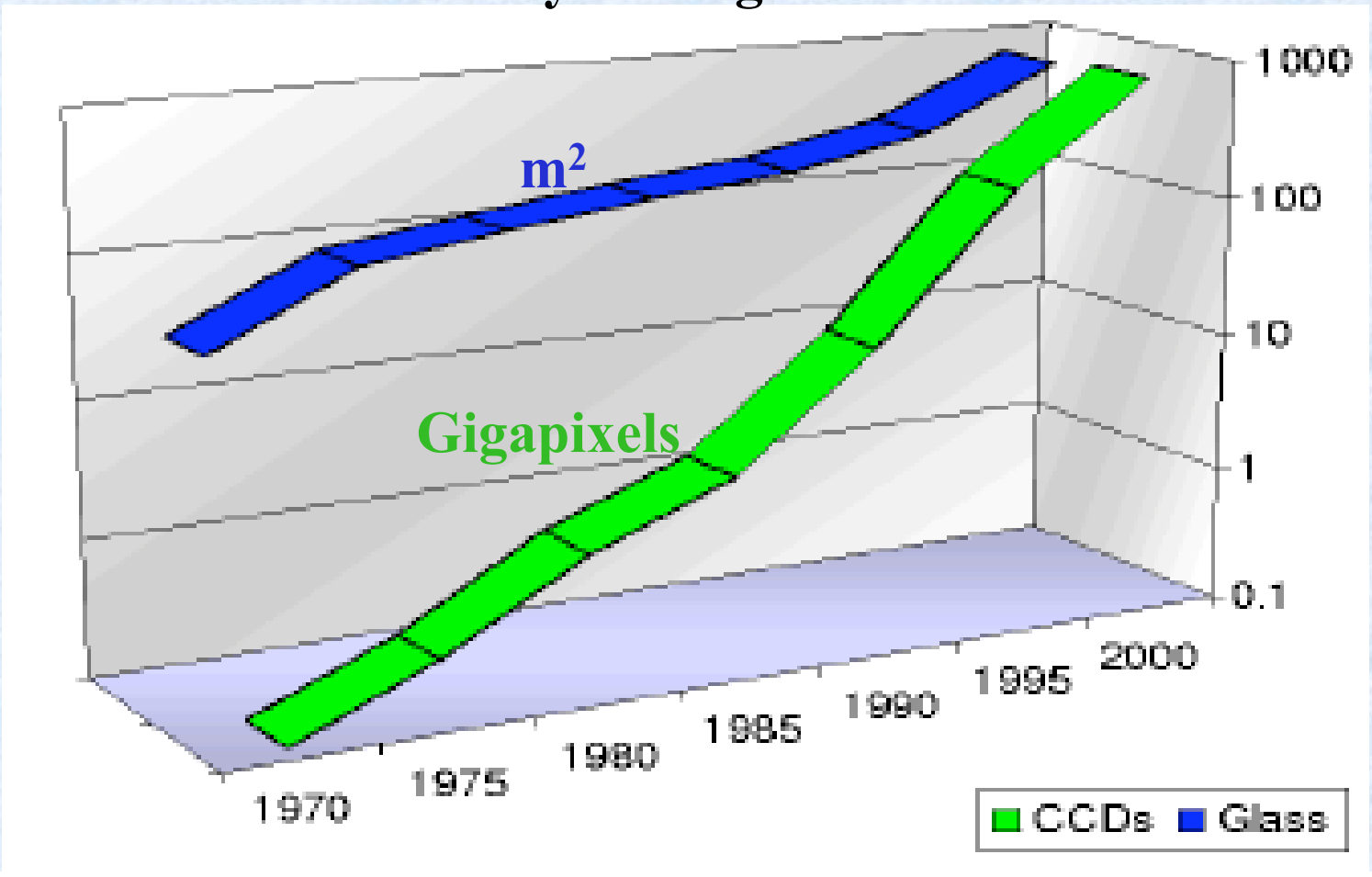
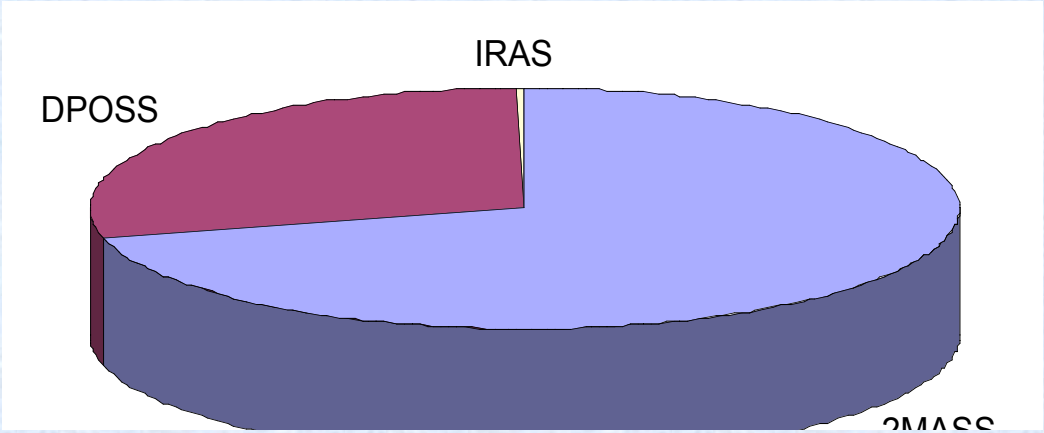


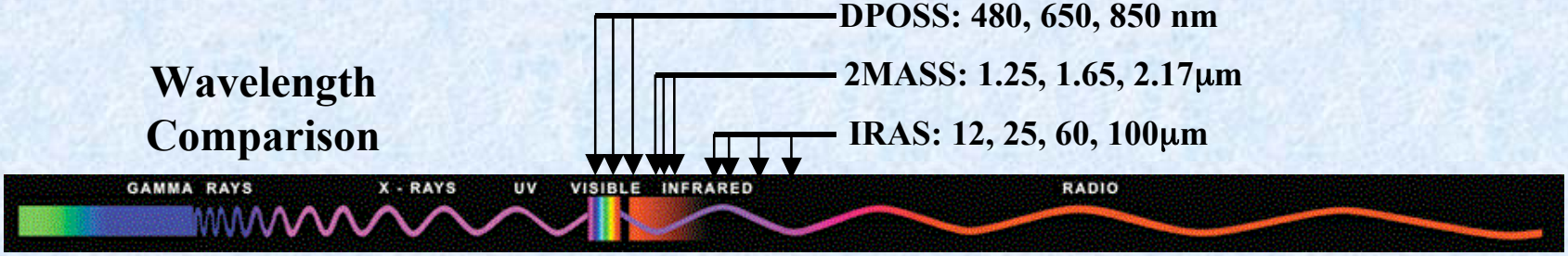
Image Archives

	IRAS	1 GB	1 arcmin	All Sky	4 Infrared Bands
	DPOSS	4 TB	1 arcsec	Northern Sky	1 Near-IR, 2 Visible Bands
	2MASS	10 TB	1 arcsec	All Sky	3 Near-Infrared Bands

Total Image Size Comparison

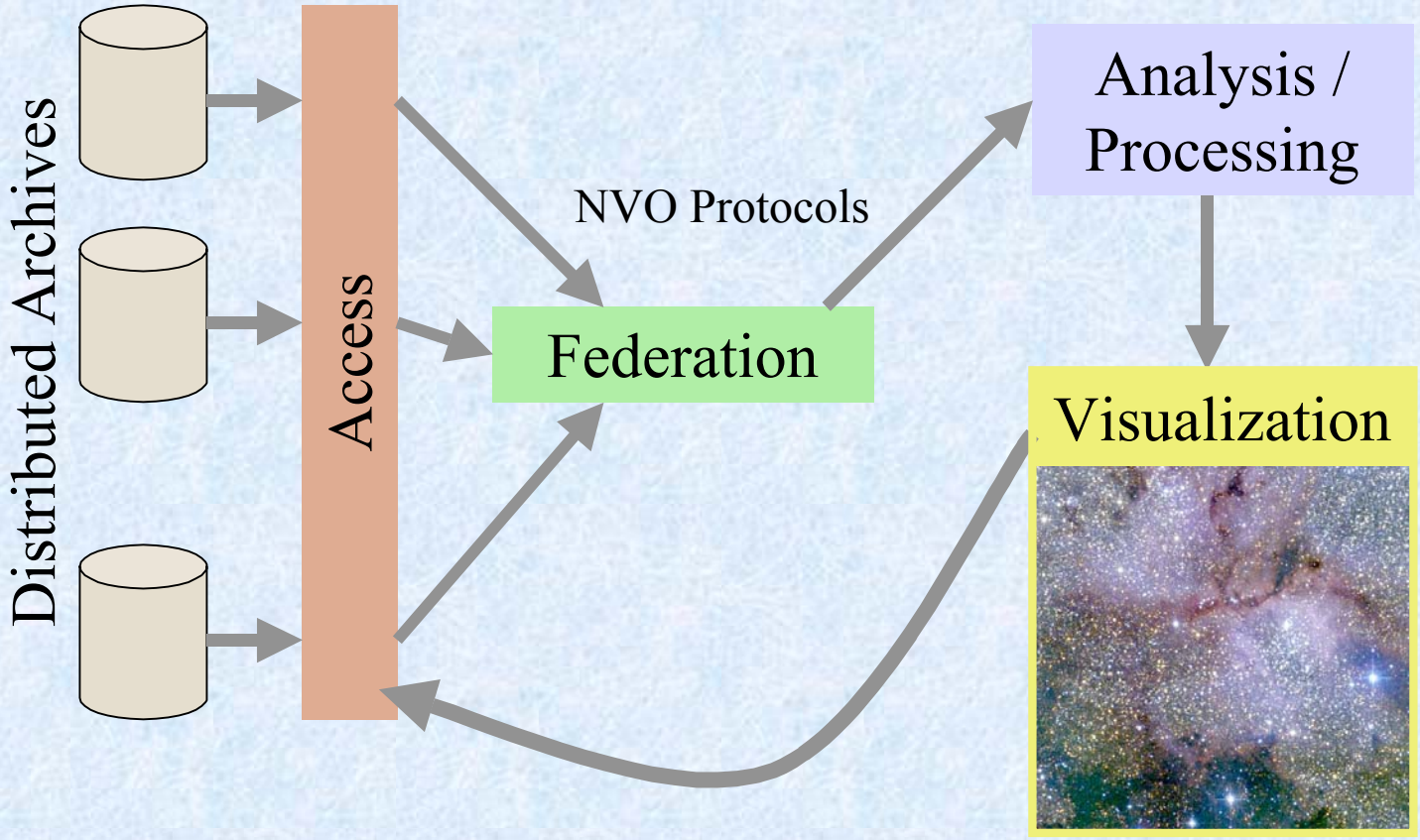


Wavelength Comparison

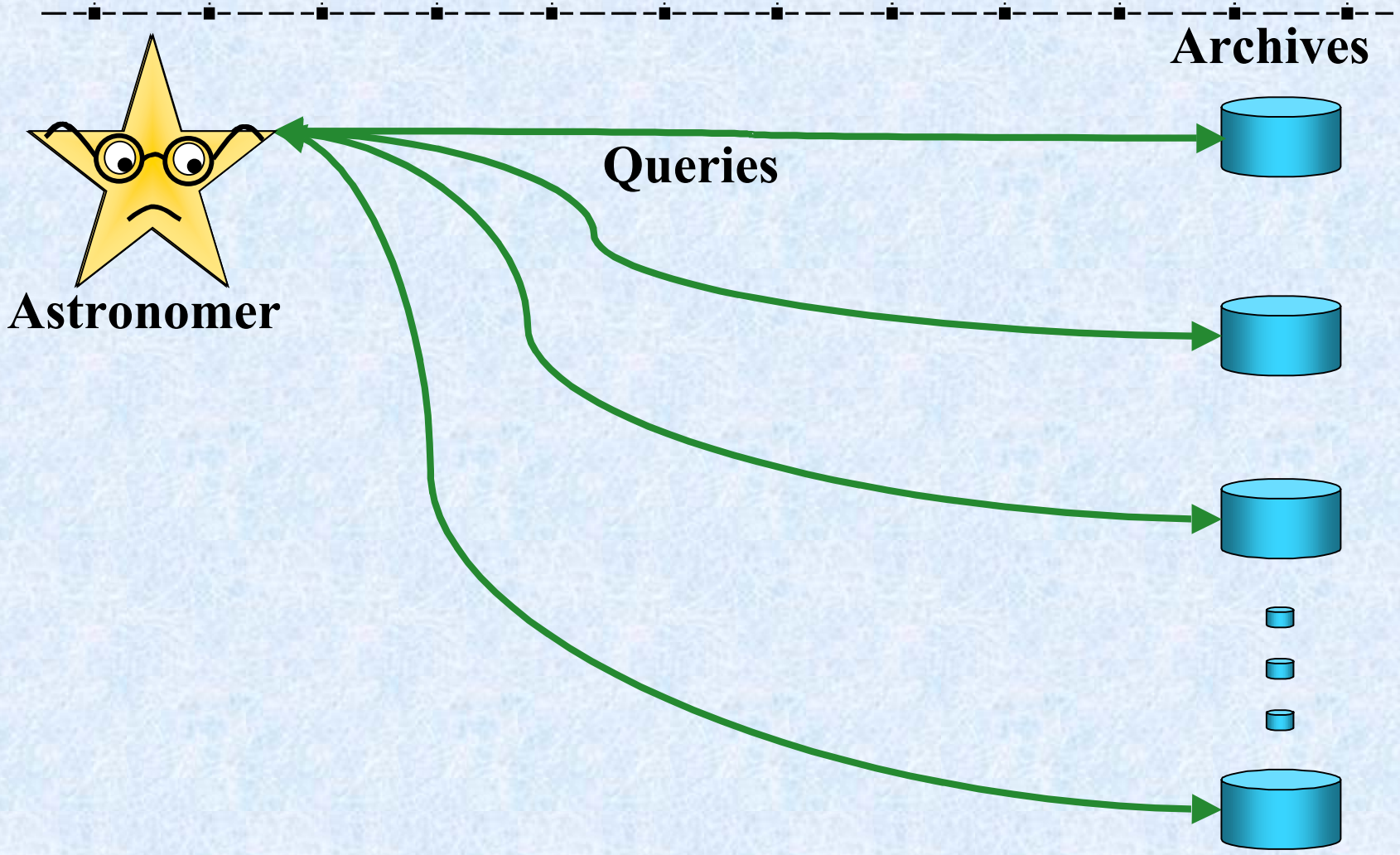


National Virtual Observatory (NVO)

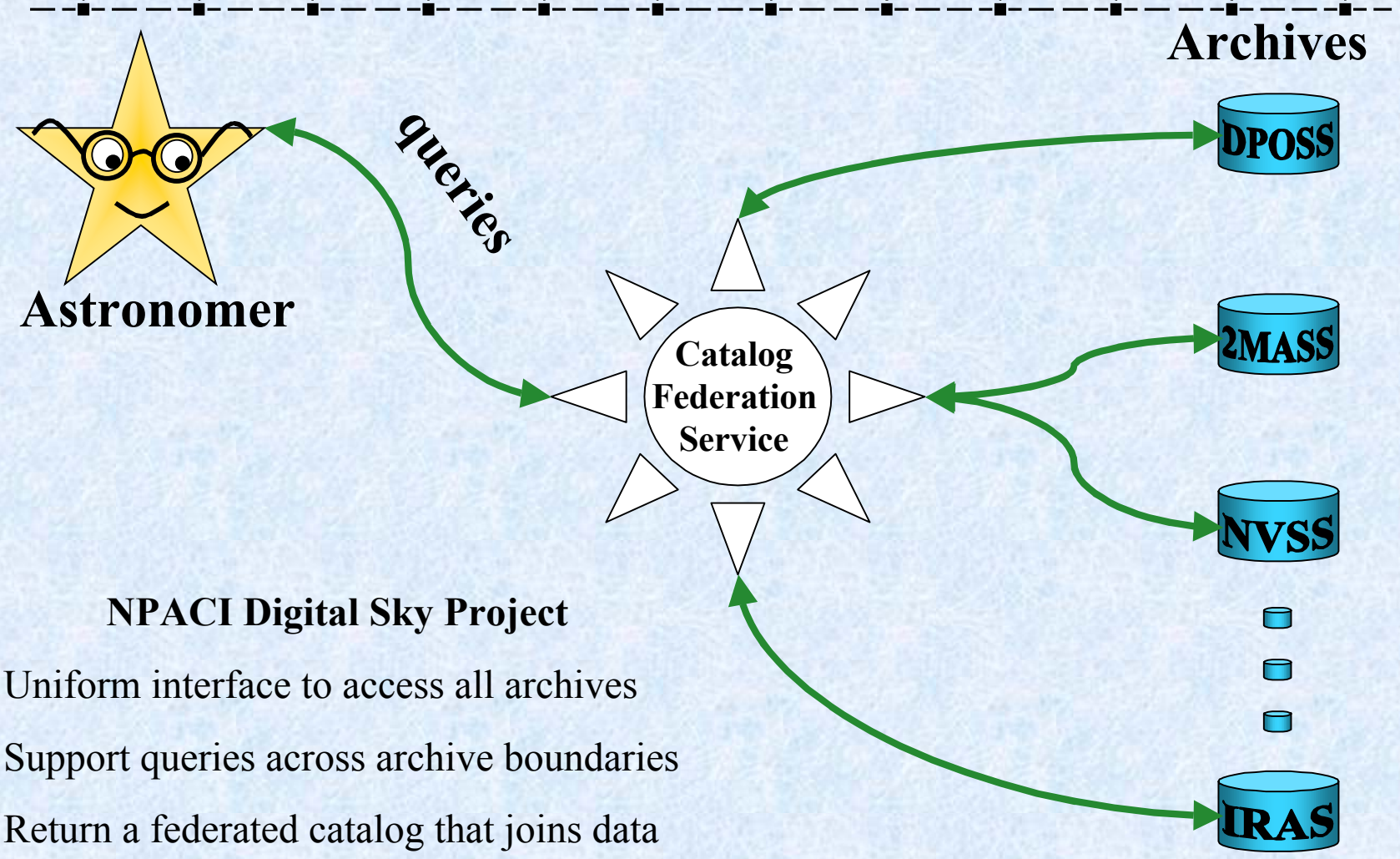
http://us-vo.org



Catalog (Metadata) Access



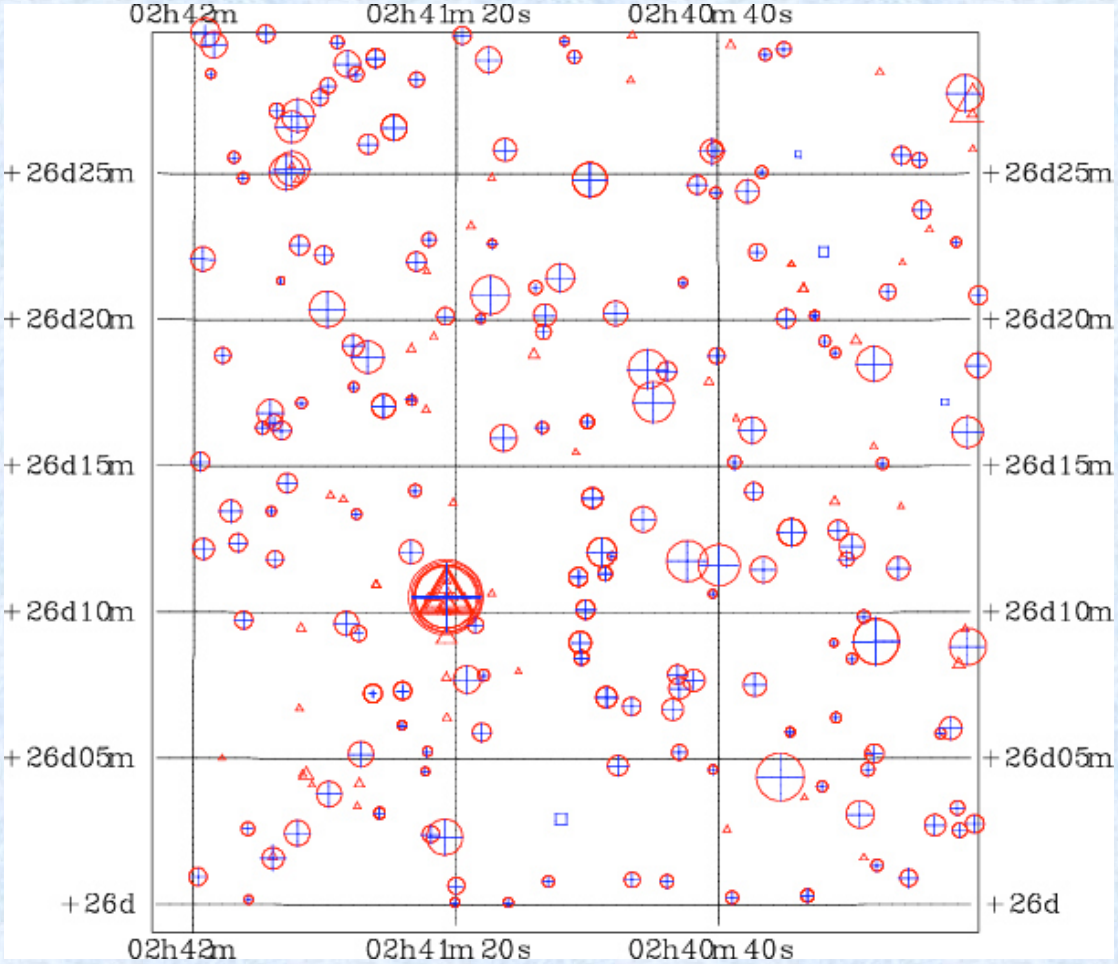
Catalog Federation Services



NPACI Digital Sky Project

- Uniform interface to access all archives
- Support queries across archive boundaries
- Return a federated catalog that joins data from multiple archives

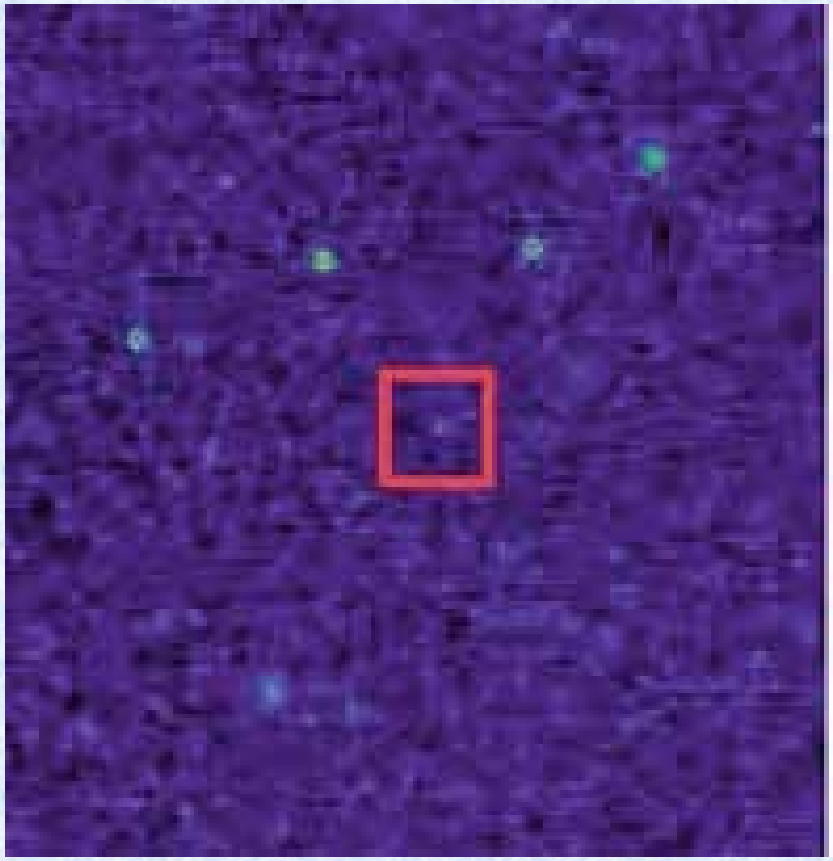
Catalog Federation



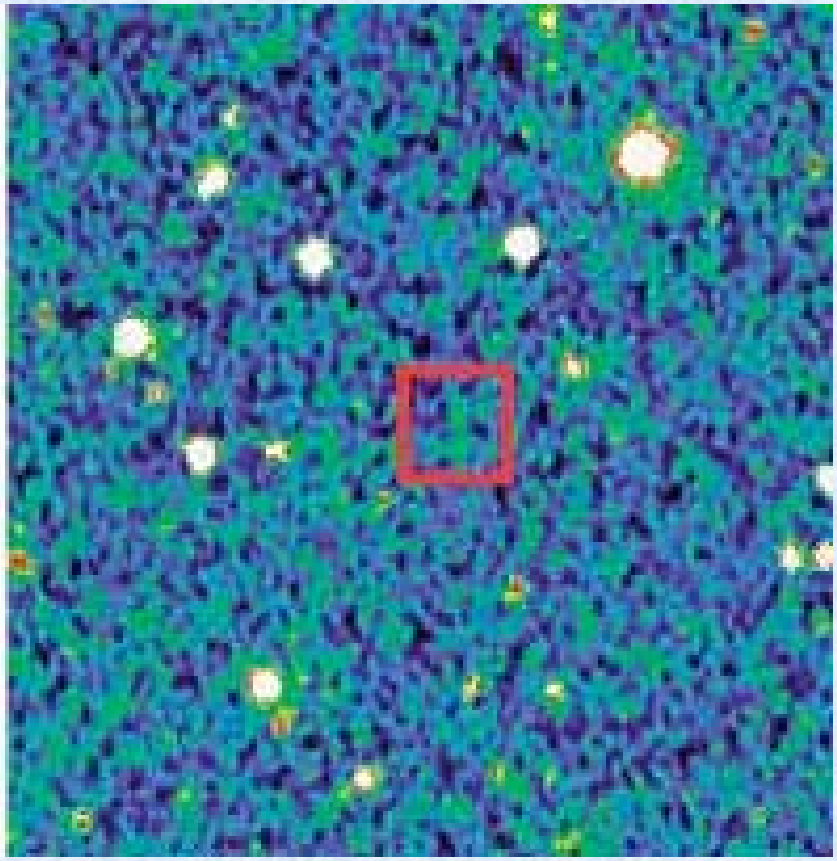
Billion Source Cross-Identification: A Computational Challenge

- SDSS unmatched
- 2MASS matched
- SDSS matched
- 2MASS unmatched

New Brown Dwarf Discovery

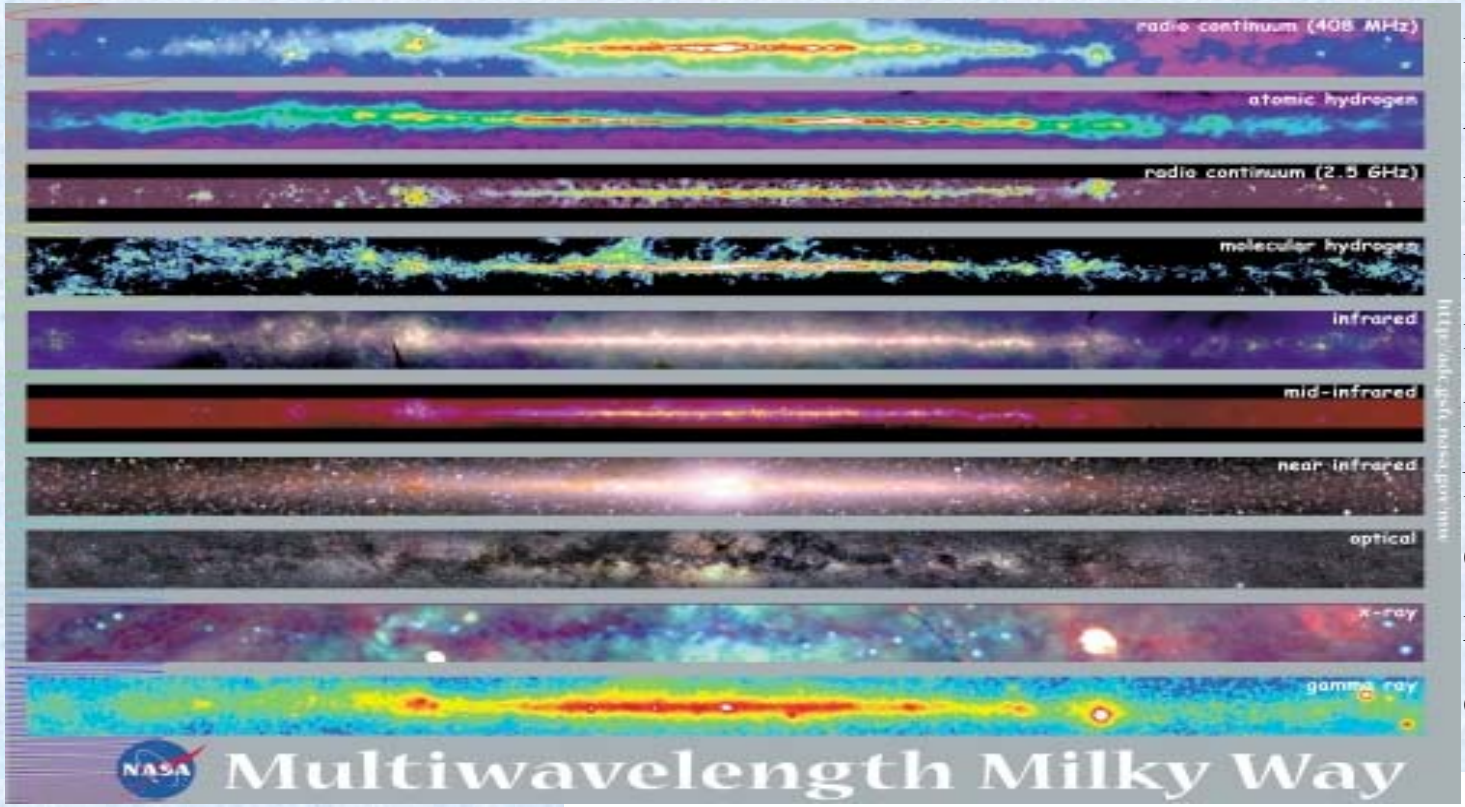


2MASS K Band (2.2 μm)



SDSS R Band (red light)

Image Federation



Radio (400 MHz)

Atomic Hydrogen

Radio (2.5 GHz)

Molecular Hydrogen

Infrared

Mid-Infrared

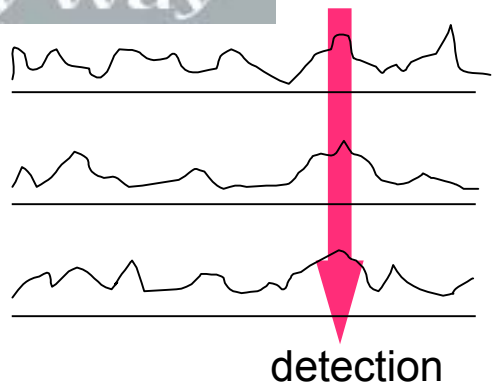
Near-Infrared

Optical

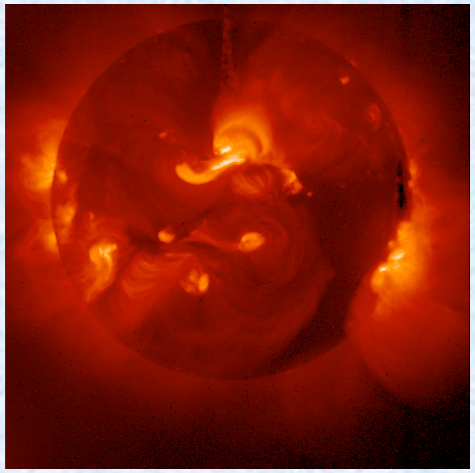
X-Ray

Gamma-Ray

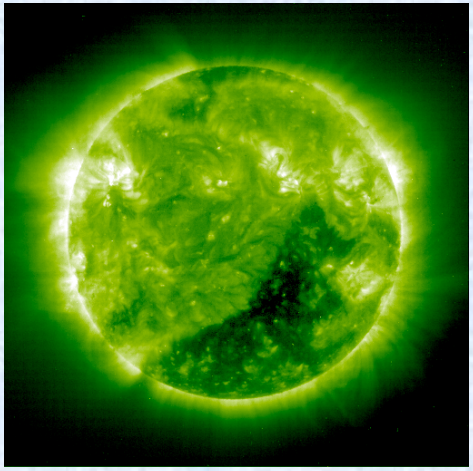
Stacking allows detection of faint sources. A 1-sigma detection in each of many bands becomes a 3-sigma detection.



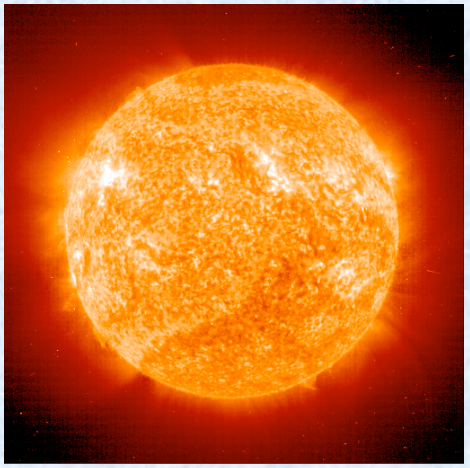
The Multi-Wavelength Sun



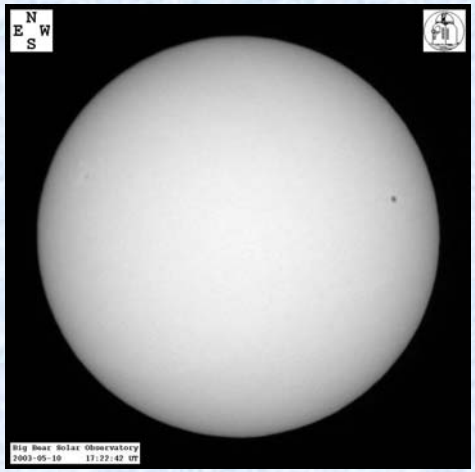
X-Ray (Yohkoh)



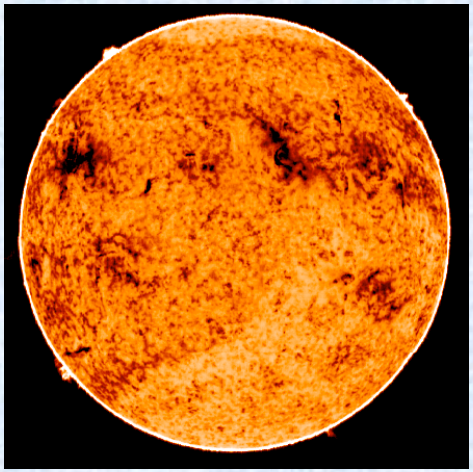
19.5 nm Ultra-Violet (SOHO)



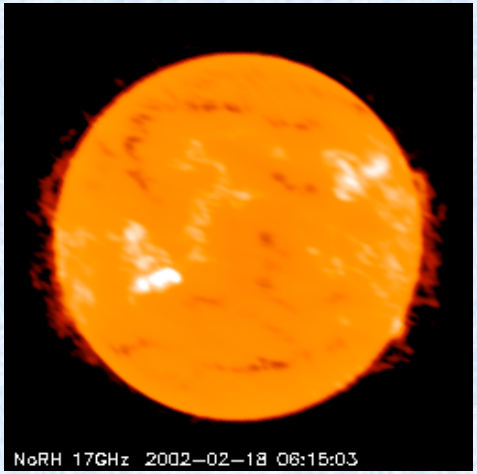
30.4 nm Ultra-Violet (SOHO)



White Light (BBSO)



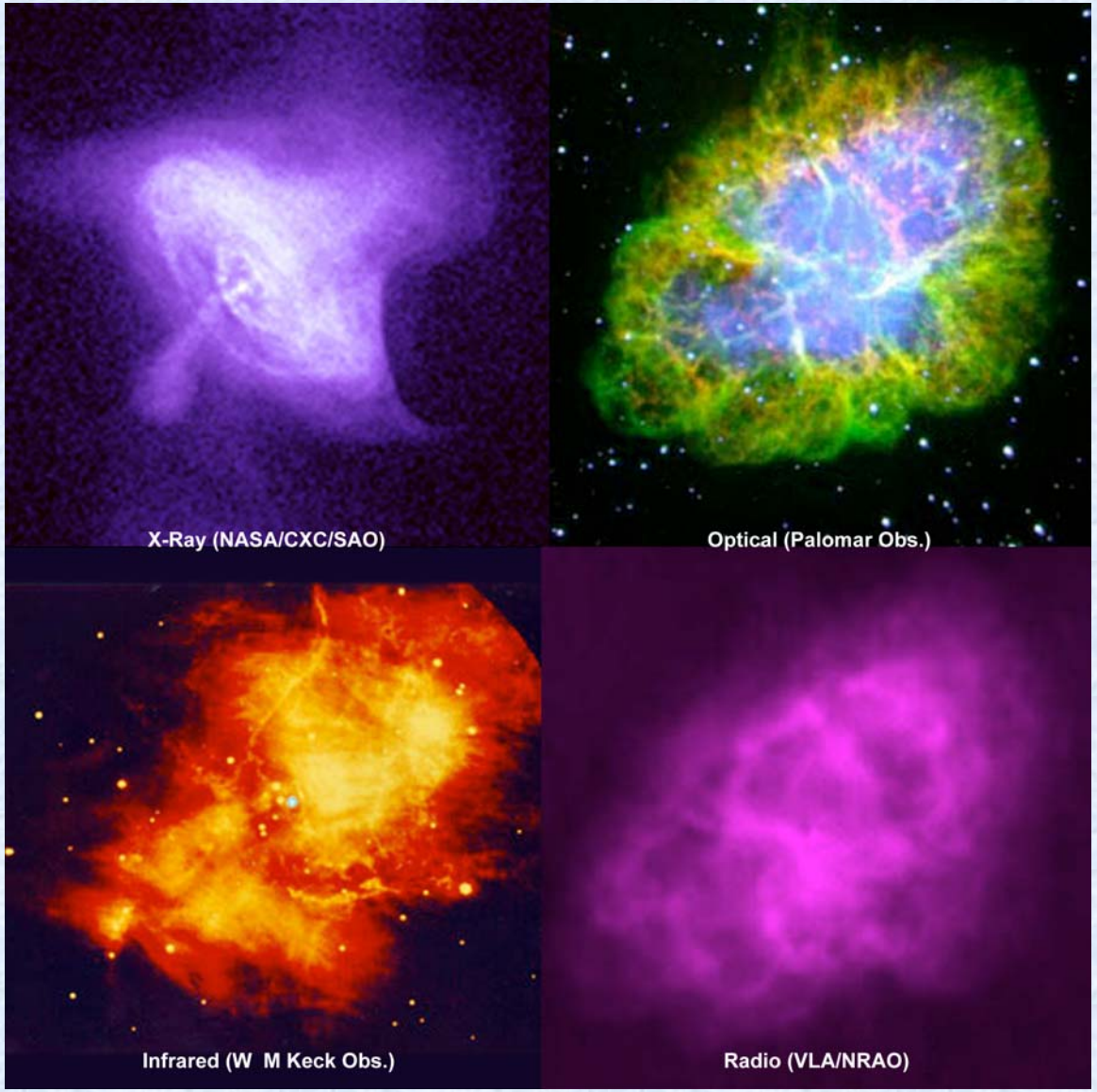
Infrared (NSO)



Radio (Nobeyama)

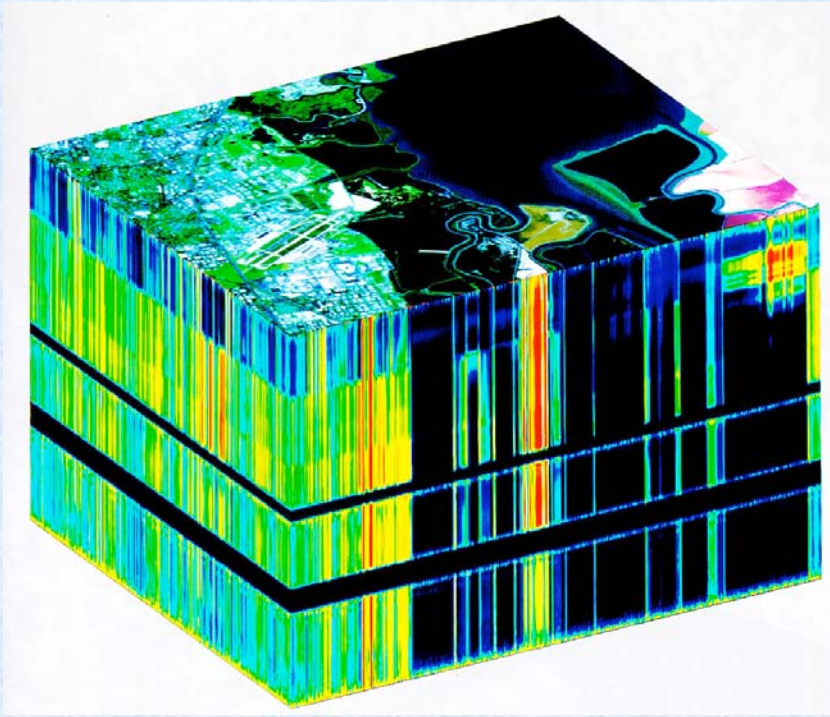
Chandra X-Ray Observatory

Crab Nebula



Hyper-Spectral Imagery

Single Instrument



Moffet Field California (AVIRIS): 224 channels from 400 nm to 2500 nm

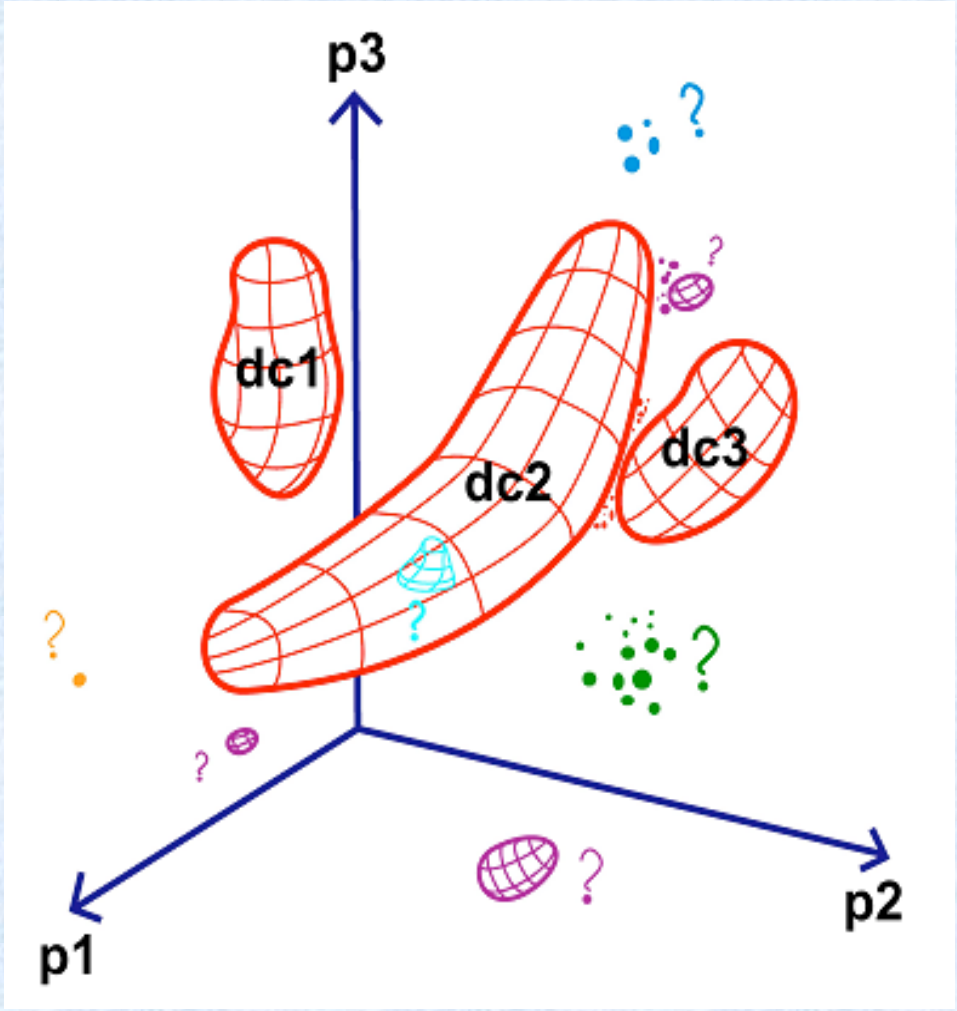
Multiple Instruments!

Crab Nebula. 3 channels: X-ray in blue, optical in green, and radio in red

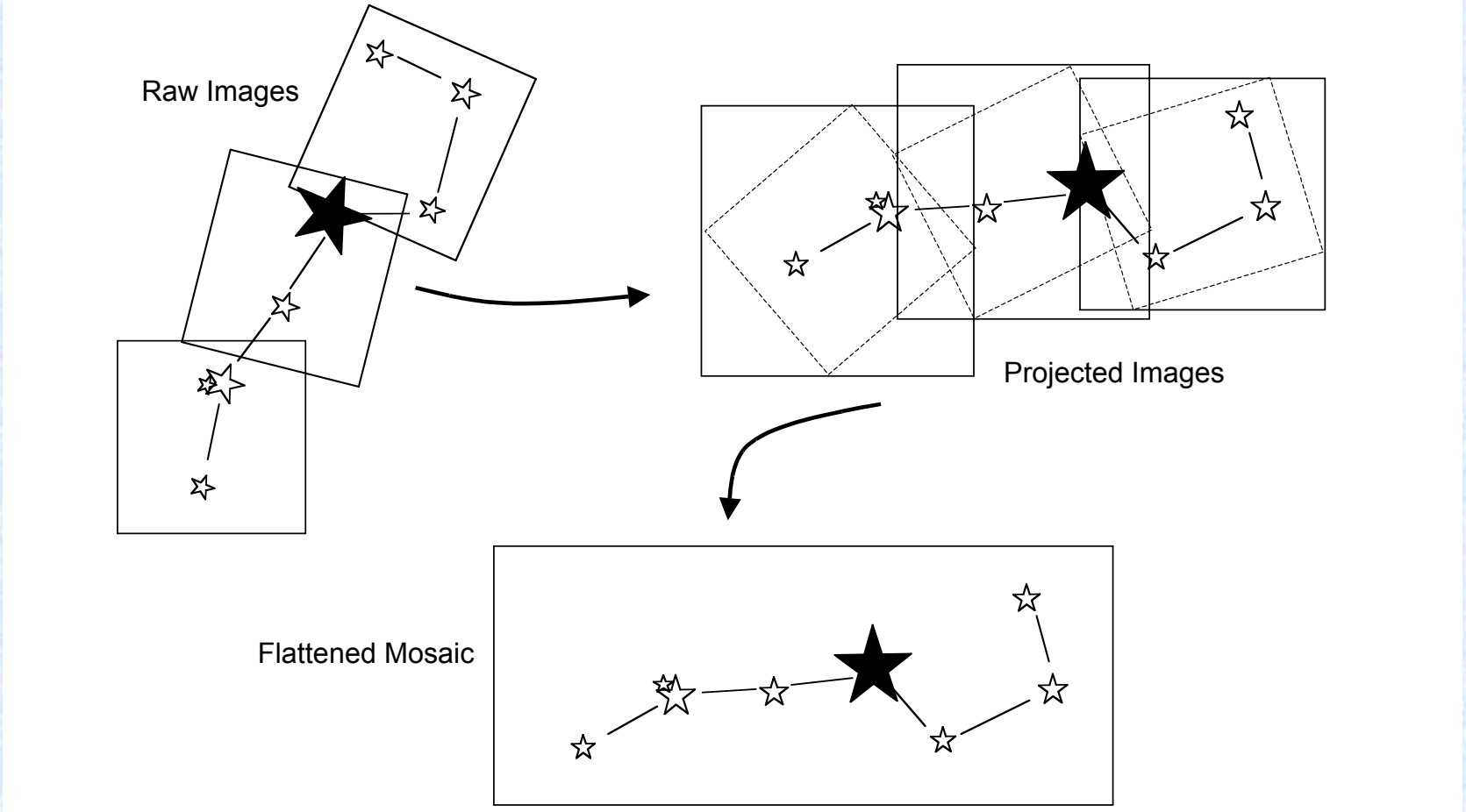
NVO Processing Example 1:
Finding Clusters and Correlations

1. Finding high-dimensional clusters and correlations
2. Finding outliers that don't cluster well

1. Discovering new objects of known type
2. Discovering entirely new types of objects!



NVO Processing Example 2:
Image Reprojection and Mosaicking



Science Drivers for Mosaics

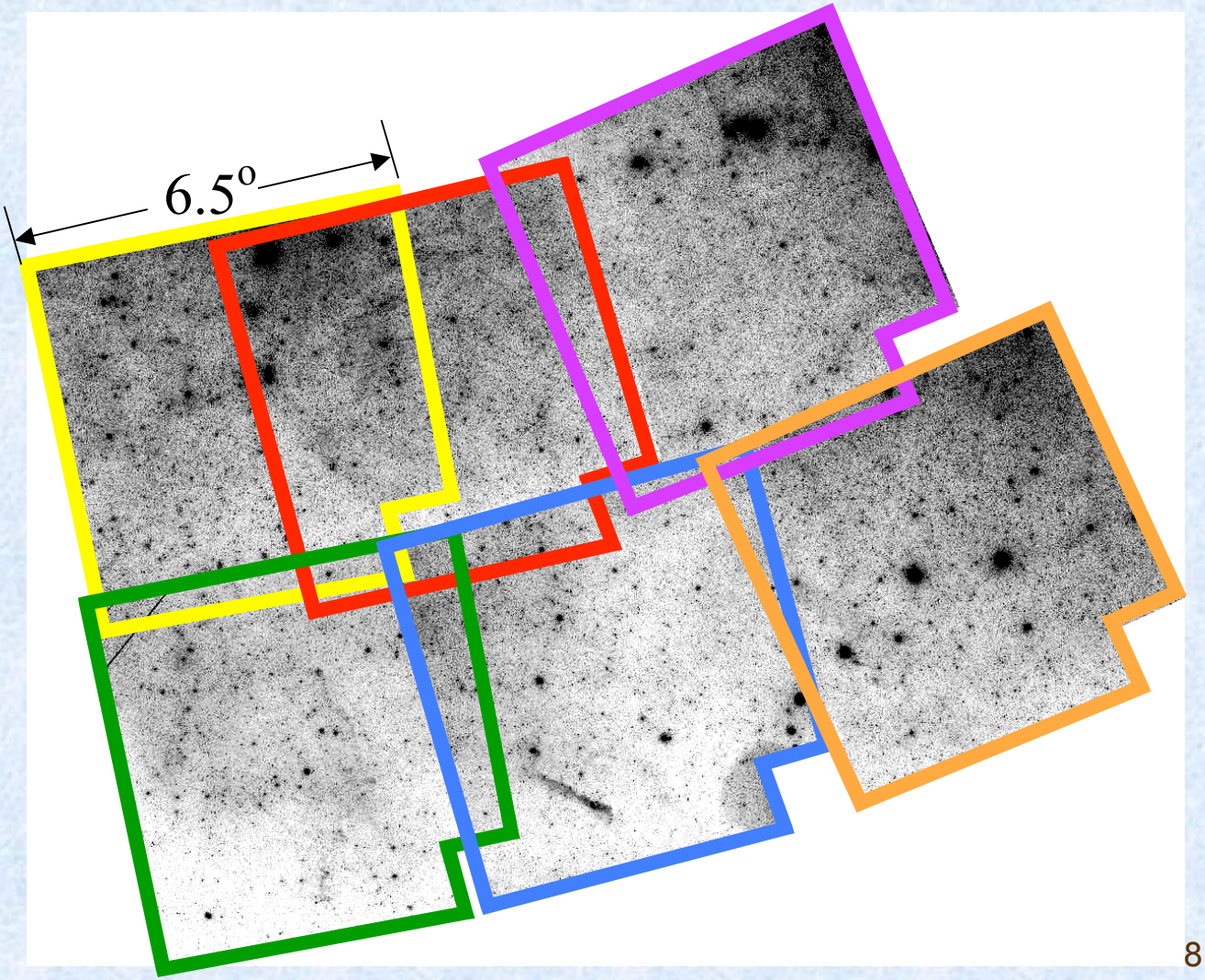
- Many important astrophysics questions involve studying regions that are at least a few degrees across.
  - Need high, uniform spatial resolution
  - BUT cameras give high resolution or wide area but not both => need mosaics
  - required for research and planning (e.g., SIRTF)
- Mosaics can reveal new structures & open new lines of research
- Star formation regions, clusters of galaxies must be studied on much larger scales to reveal structure and dynamics
- Mosaicking multiple surveys to the same grid – **image federation** – required to effectively search for faint, unusual objects, transients, or unknown objects with unusual spectrum.

Mosaics Enable Study of Large Scale Structure

High Velocity Clouds (HVC) can span many DPOSS plates.

HVCs cover over a third of the sky!

By comparison, Earth's full moon spans about _ degree when viewed from Earth.



Reprojection to a Common Grid Enables Multi-Wavelength Science:

Birth of a Globular Cluster?

30 Doradus: a hotbed of star formation, supernova explosions, and ionized plasma in the Large Magellanic Cloud. Astronomers expect that a globular cluster will have formed here in a few million years.

Red: X-Ray emission from super-hot gases.

Green: Emission from ionized hydrogen gas.

Blue: Ultra-violet radiation from hot stars.

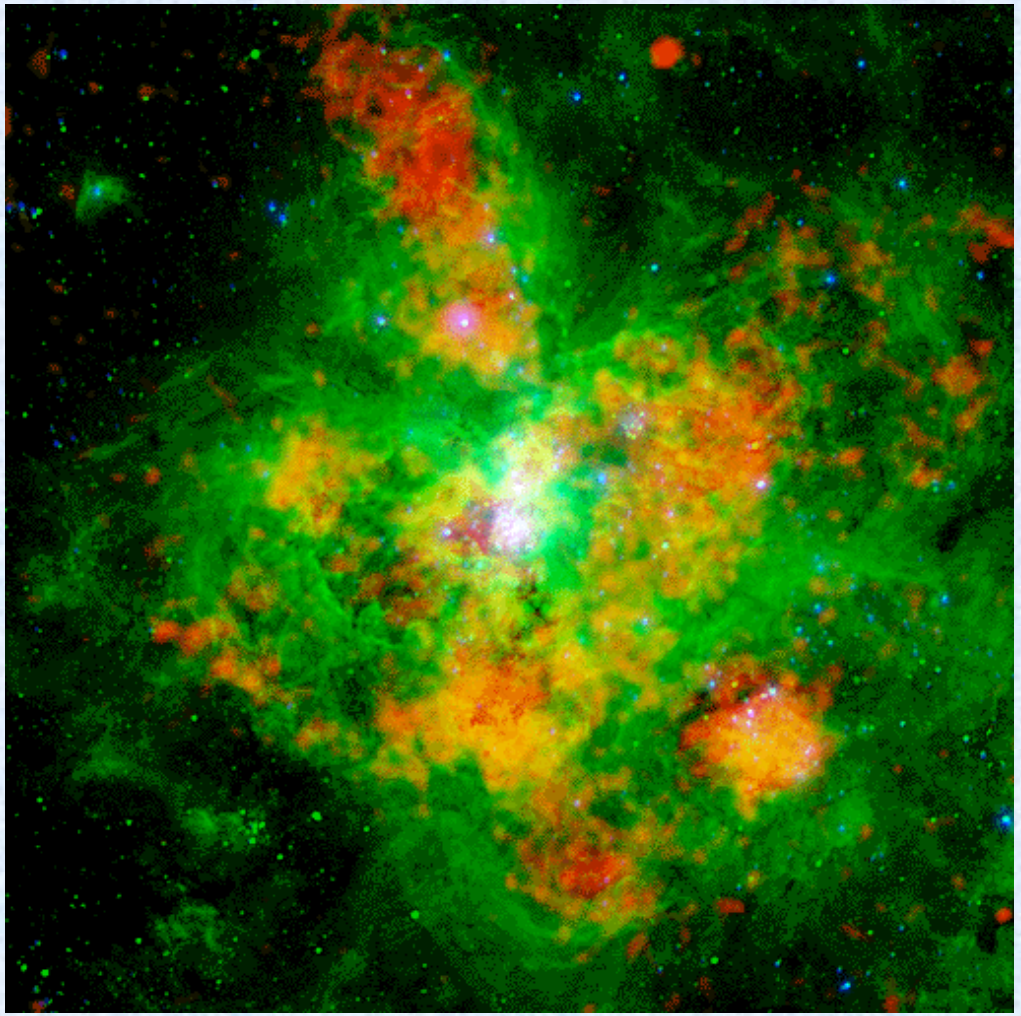


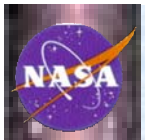
Image Credit: Q. Daniel Wang (NWU), UM/CTIO, UIT, ROSAT, http://apod.gsfc.nasa.gov

NVO Interoperability

- Use Web Services Standards
  - XML: eXtensible Markup Language
  - SOAP: Simple Object Access Protocol
  - WSDL: Web Services Description Languages
- Define standards for astronomical data
  - VOTable for catalogs
  - FITS for images encapsulated with metadata
- Define standards for serving data
  - Cone Search
  - Simple Image Access Protocol (SIAP)

Virtual Observatories

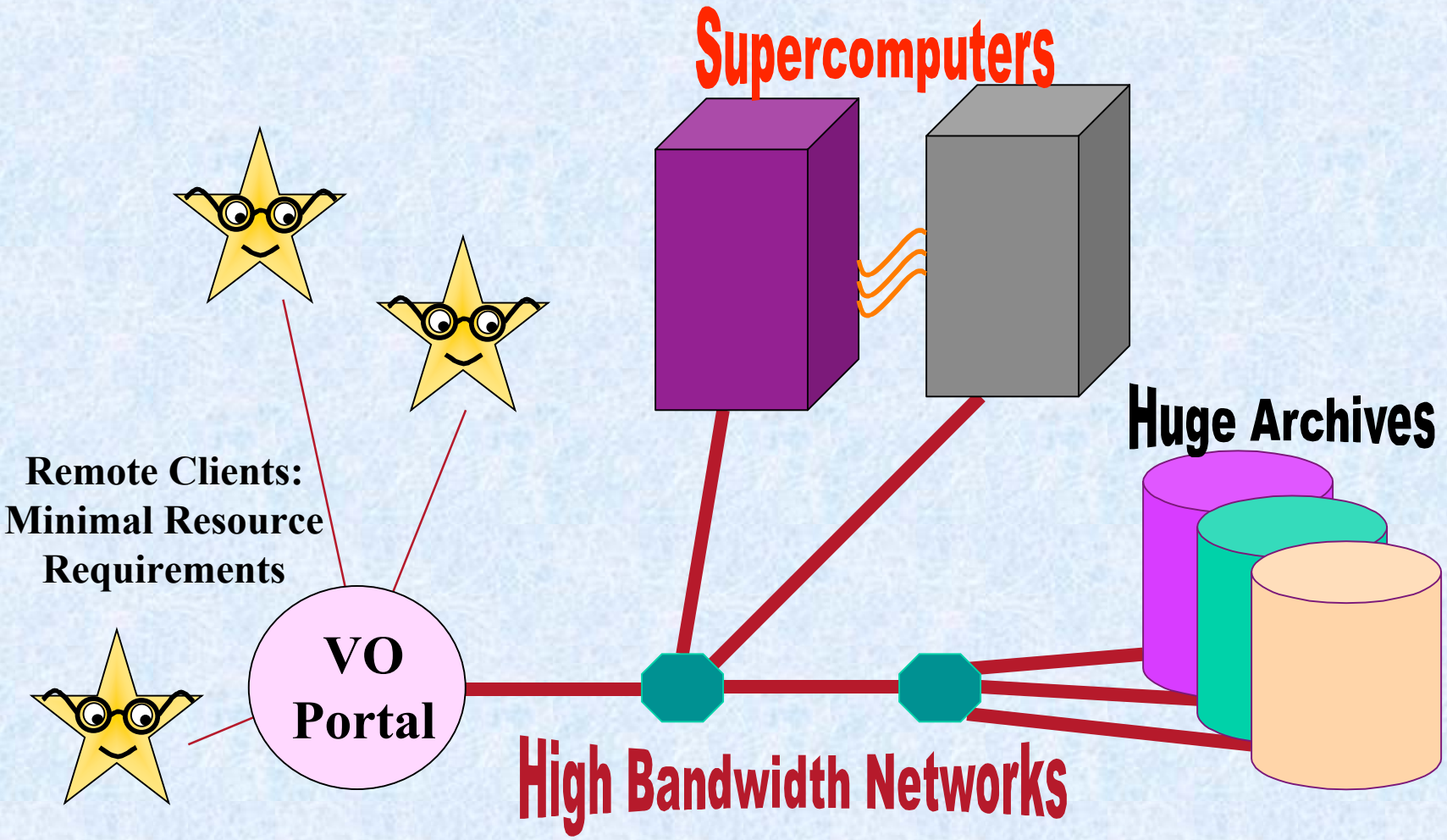
- Community driven.
- Community built.
- Community accessible.
- Emphasis on many interoperable components developed and deployed by domain experts in different areas.
- Highly distributed, including centers for:
  - Archive
  - Processing
  - Visualization
- Exploit high performance computation and communications resources.



Outline

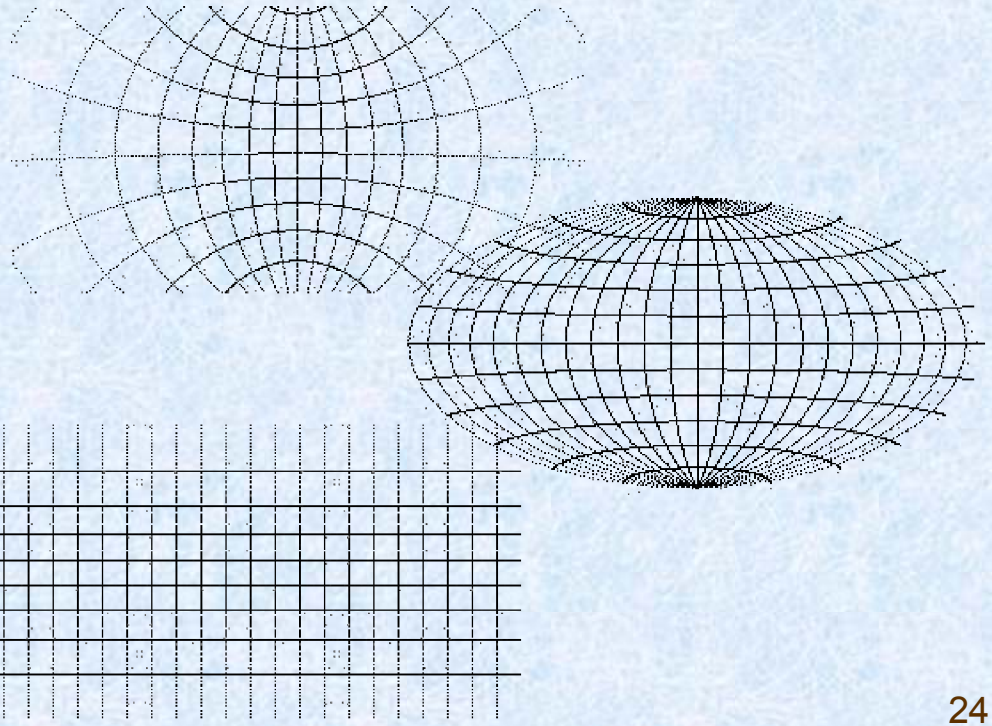
- Virtual Observatories
  - Motivation
  - National Virtual Observatory
- **yourSky Custom Mosaic Server**
- Montage: High Science Quality Mosaics
- Visualization
  - High Performance / Powerwall
  - Web-Based / Desktop Access
- Computational Grids

Transparent Use of High Performance Infrastructure



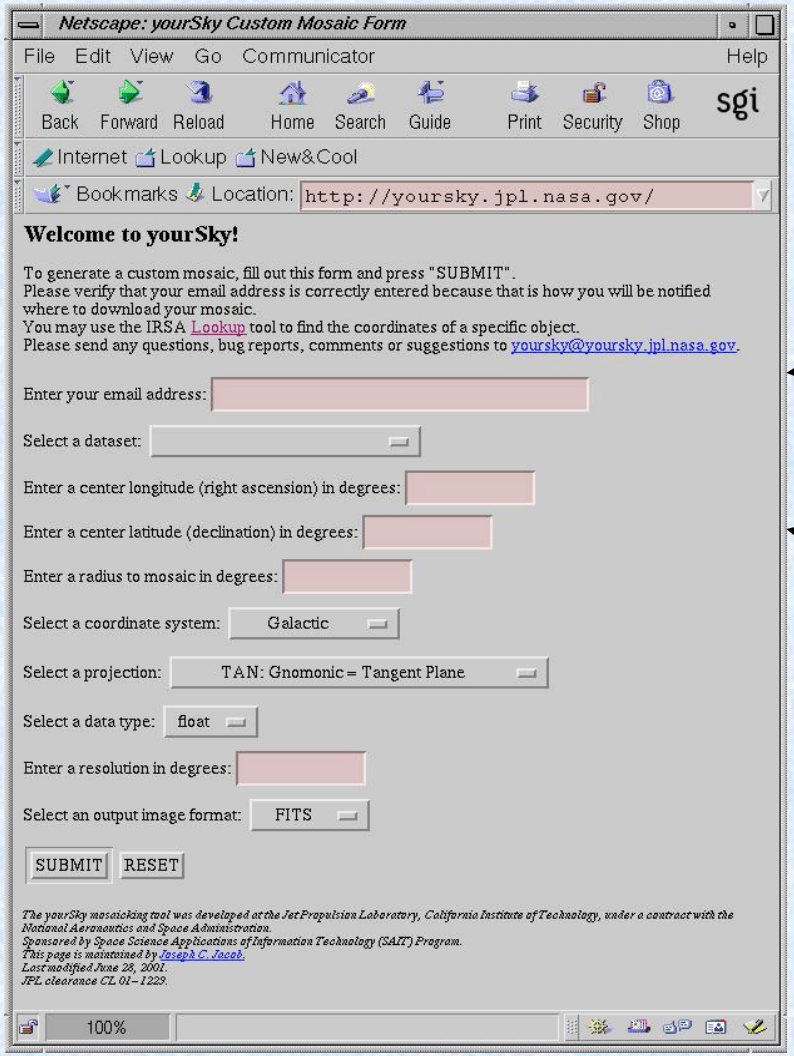
Example: Custom 'On-the-fly' Mosaic Portal

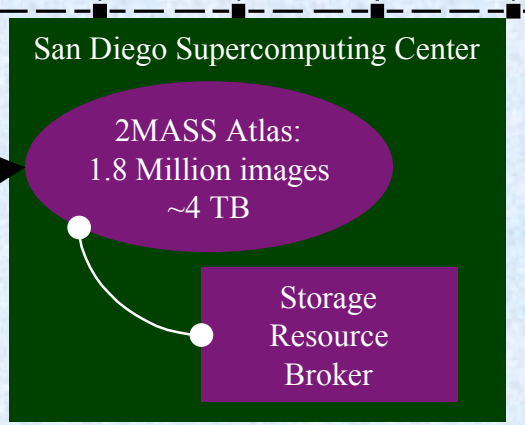
- Provide **custom access** to a compute-intensive, scalable, interoperable service that delivers science-grade image mosaics to user's desktops, through existing portals.
- Custom access = user specifies parameters
  - Data set to be used
  - Location on the sky
  - Mosaic size
  - Pixel resolution
  - Coordinate system
  - Projection
  - Data type
  - Image format



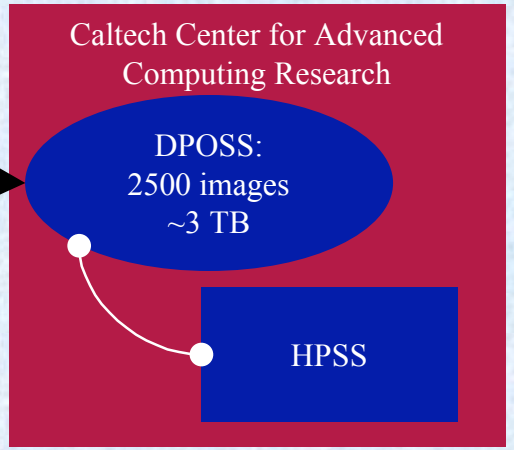
yourSky Custom Mosaic Portal

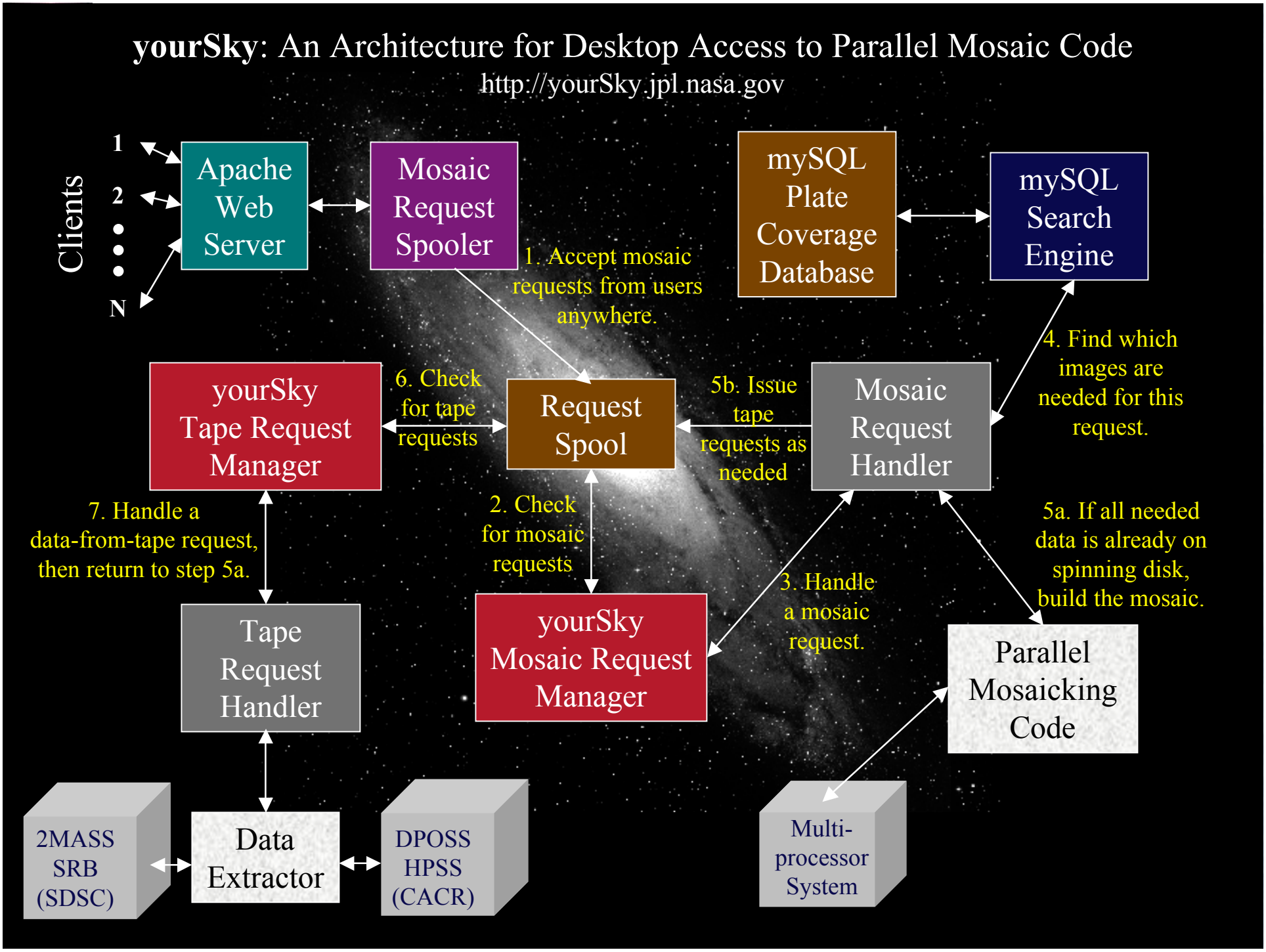
http://yourSky.jpl.nasa.gov





yourSky can access all of the publicly released DPOSS and 2MASS images for custom mosaic construction.

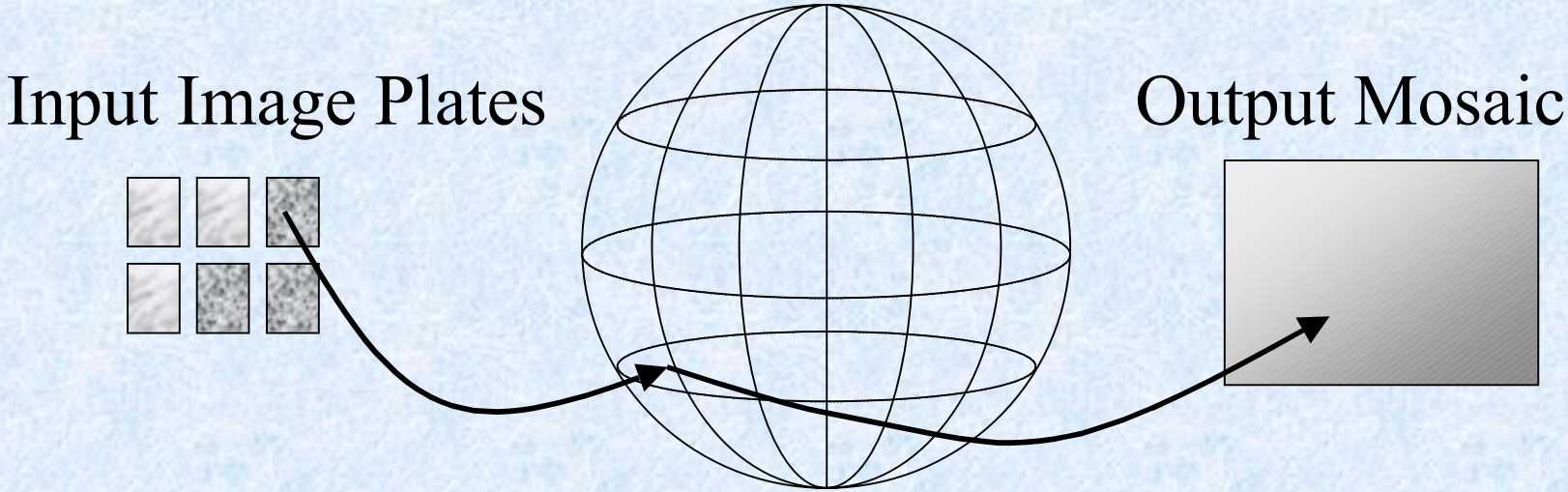



yourSky: An Architecture for Desktop Access to Parallel Mosaic Code
http://yourSky.jpl.nasa.gov
Clients
1
2
N
Apache Web Server
Mosaic Request Spooler
1. Accept mosaic requests from users anywhere.
mySQL Plate Coverage Database
mySQL Search Engine
4. Find which images are needed for this request.
yourSky Tape Request Manager
6. Check for tape requests
Request Spool
5b. Issue tape requests as needed
Mosaic Request Handler
7. Handle a data-from-tape request, then return to step 5a.
2. Check for mosaic requests
5a. If all needed data is already on spinning disk, build the mosaic.
3. Handle a mosaic request.
Tape Request Handler
yourSky Mosaic Request Manager
Parallel Mosaicking Code
2MASS SRB (SDSC)
Data Extractor
DPOSS HPSS (CACR)
Multi-processor System

yourSky Components

- Parallel Image Mosaicking Code
- Request Management
- Plate Coverage Database
- Data Management
  - Retrieval of image plates from remote archives
  - Local data cache

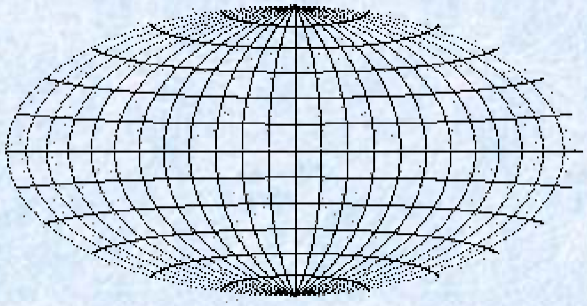
Mosaicking Code – Coordinate Mapping



1. Map input pixel coordinates (x_i,y_i) to sky coordinates (RA,Dec).

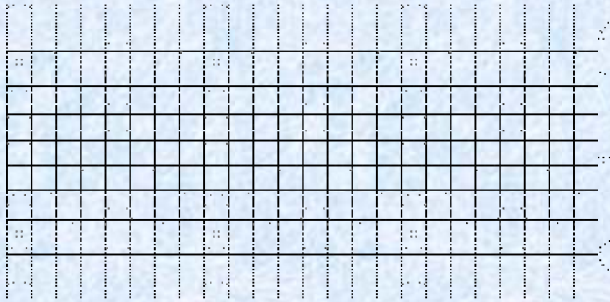
2. Map sky coordinates (RA,Dec) to output pixel coordinates (x_o,y_o).

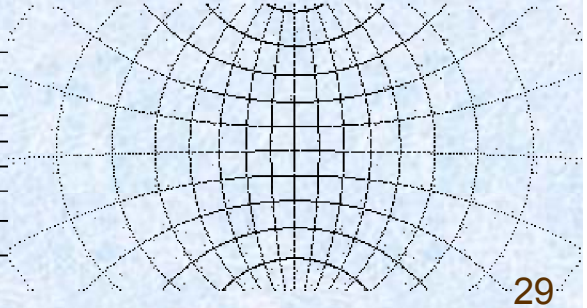
$(x_i,y_i) \longrightarrow (RA,Dec) \longrightarrow (x_o,y_o)$



Custom Coordinate System and Projection

- Coordinate Systems: Galactic, Ecliptic, J2000 Equatorial, B1950 Equatorial.
- WCS projections: LIN, TAN, SIN, STG, AZP, ARC, ZPN, ZEA, AIR, CYP, CAR, MER, CEA, COP, COD, COE, COO, BON, PCO, SFL, PAR, AIT, MOL, CSC, TSC, DSS, PLT.





Custom Image Format and Data Type

Image Format:

- FITS
- JPEG
- PGM
- PNG
- TIFF
- Raw Data

Data Type:

- 8-bit unsigned integer
- 8-bit signed integer
- 16-bit unsigned integer
- 16-bit signed integer
- 32-bit unsigned integer
- 32-bit signed integer
- Single precision floating point
- Double precision floating point

Request Management

- User's priority is reduced according to the number of pixels generated in past 24 hours.
- A user that has had to wait for data retrieval from tape gets high priority once the data is ready.
- Ensures:
  - All users get a chance to have their mosaic processed.
  - No single user dominates the available resources.

Plate Coverage Database

- Contains the minima and maxima of the longitudes (RAs) and latitudes (declinations) in each supported coordinate system.
- Query result is a list of the input image plates required to fulfill the mosaic request.
- Accessible as a stand-alone service: http://yourSky.jpl.nasa.gov/query/index.html.

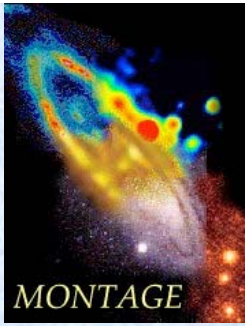
Data Management

- Input image plates used to construct mosaics are cached locally for later requests and discarded on a least recently used basis.
- Completed mosaics are stored in a work area for 24 hours (for user download).
- Input image plates are retrieved from:
  - **DPOSS:** High Performance Storage System (HPSS) at Caltech's Center for Advanced Computing Research.
  - **2MASS:** Storage Resource Broker (SRB) at the San Diego Supercomputer Center.

Outline

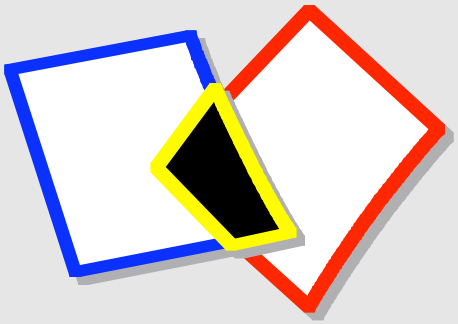
- Virtual Observatories
  - Motivation
  - National Virtual Observatory
- yourSky Custom Mosaic Server
- **Montage: High Science Quality Mosaics**
- Visualization
  - High Performance / Powerwall
  - Web-Based / Desktop Access
- Computational Grids

From yourSky to Montage

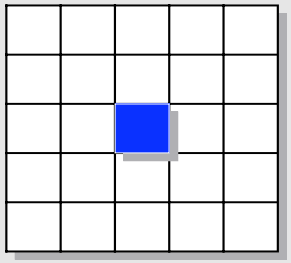


- yourSky is the baseline code for Montage, an Earth Science Technology Office – Computational Technologies Project Round 3 Grand Challenge.
- Collaboration between CACR, IPAC and JPL.
- Montage will improve upon yourSky mosaicking code:
  - **Science Quality** – Flux preservation / Background matching.
  - **Interoperability** with NVO infrastructure.
  - **Interoperability** with TeraGrid infrastructure.
- Montage has staged code improvement deliverables through January, 2005.
- http://montage.ipac.caltech.edu

Montage Reprojection Module



Arbitrary Input Image



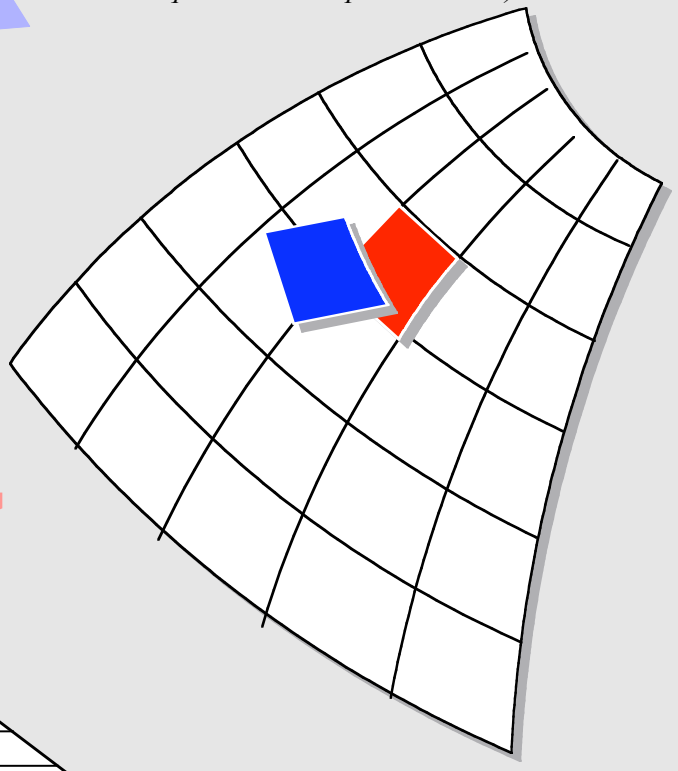
Central to the algorithm is accurate calculation of the area of spherical polygon intersection between two pixels (assumes great circle segments are adequate between pixel vertices)

Input pixels projected on celestial sphere

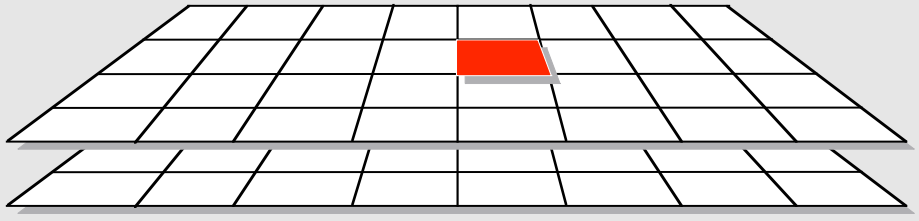
```
SIMPLE  =                    T /
 BITPIX  =                  -64 /
 NAXIS   =                    2 /
 NAXIS1  =                 3000 /
 NAXIS2  =                 3000 /
 CDELT1  =       -      3.333333E-4 /
 CDELT2  =              3.333333E-4 /
 CRPIX1  =               1500.5 /
 CRPIX2  =               1500.5 /
 CTYPE1  = 'RA---TAN'           /
 CTYPE2  = 'DEC--TAN'           /
 CRVAL1  =           265.91334 /
 CRVAL2  =           -29.35778 /
 CROTA2  =                  0. /
END
```

FITS header defines output projection

Output pixels projected on celestial sphere



Reprojected Image

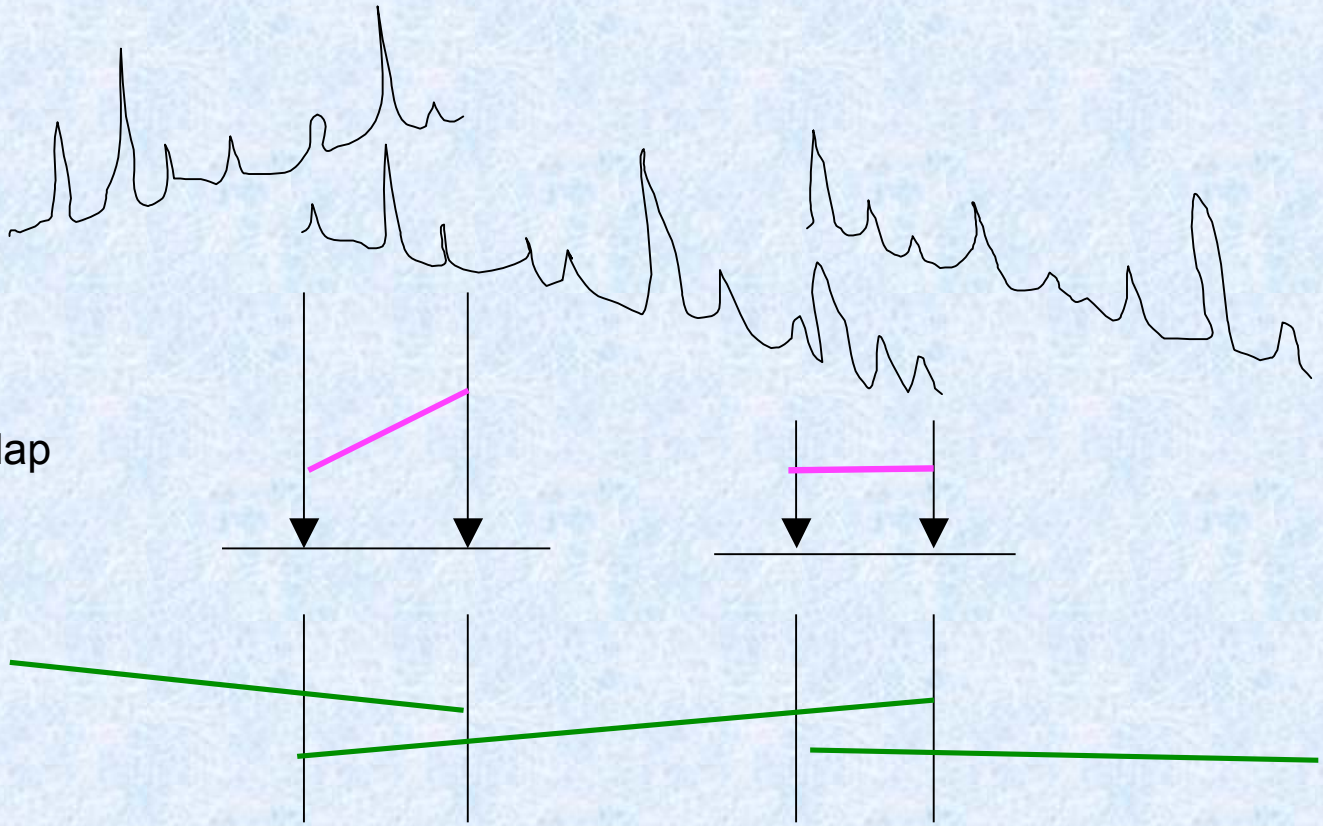


Total Flux

Sky Area Coverage (steradians)

Montage Background Matching

Project pixels to output projection



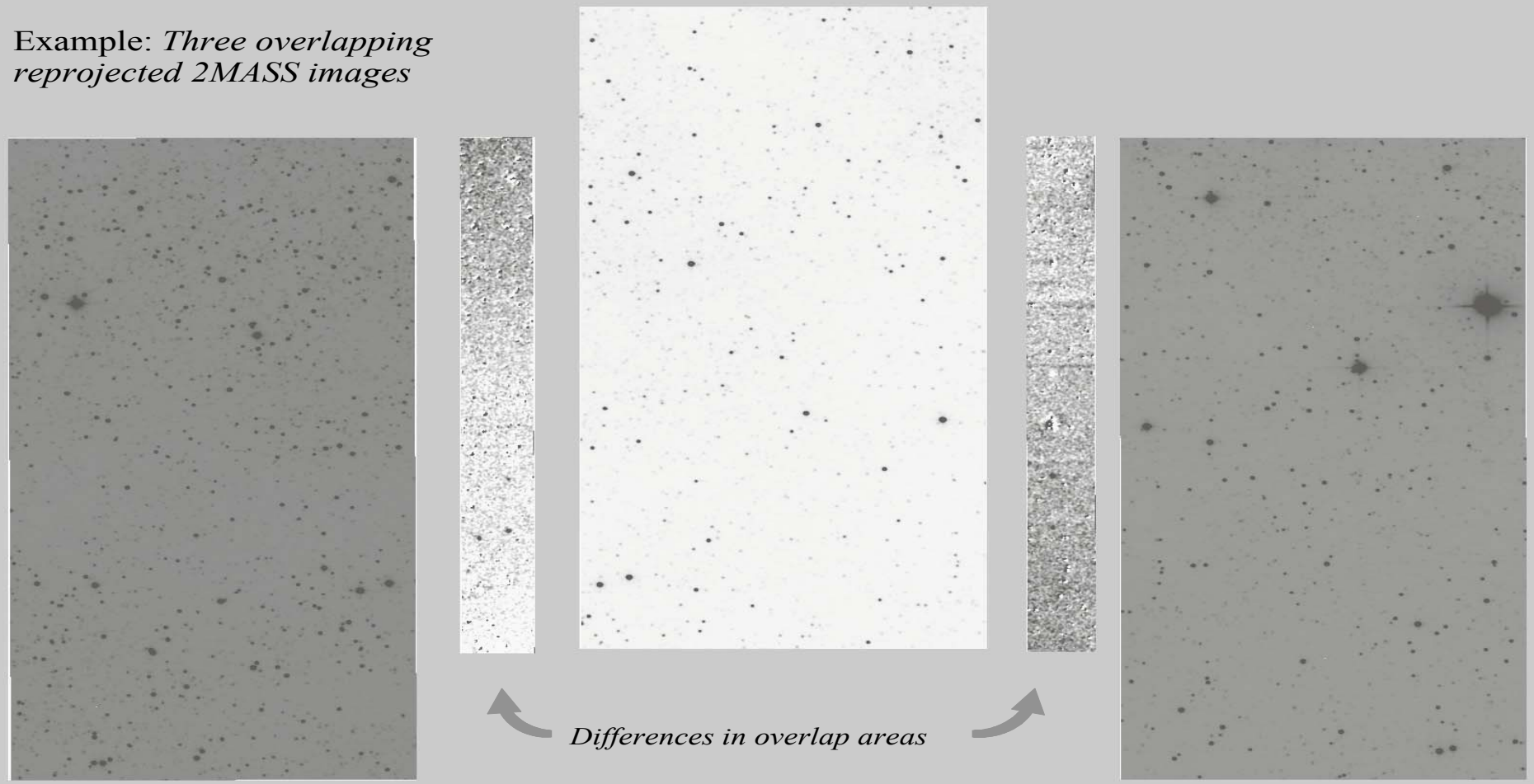
Fit ramps to overlap differences

Find correction ramp for each image.

Subtract from Pixel values

Montage Background Correction Procedure

Example: *Three overlapping reprojected 2MASS images*

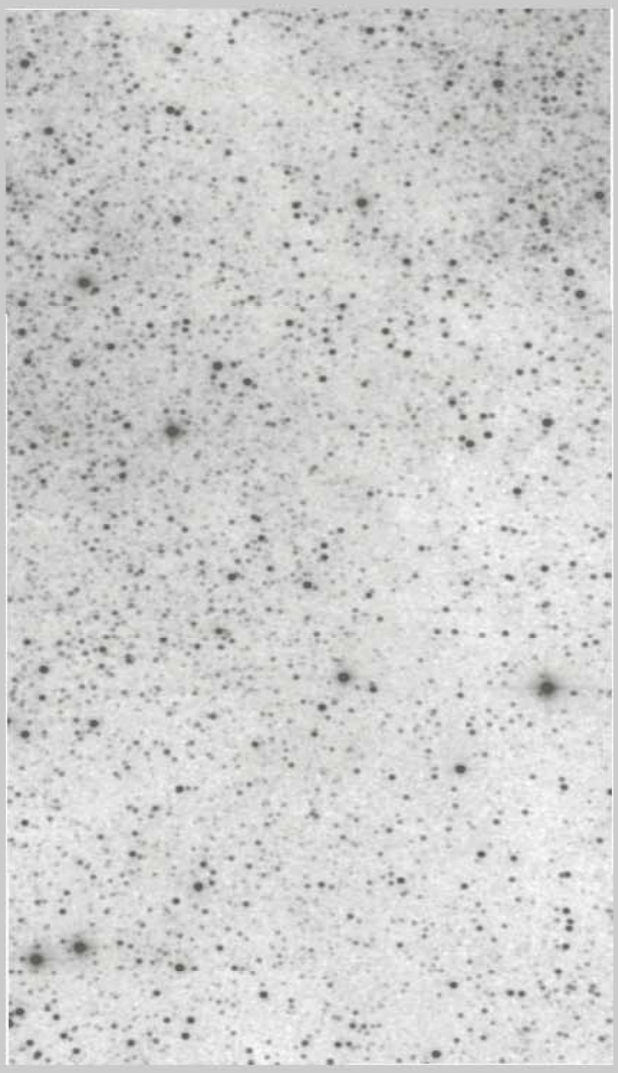


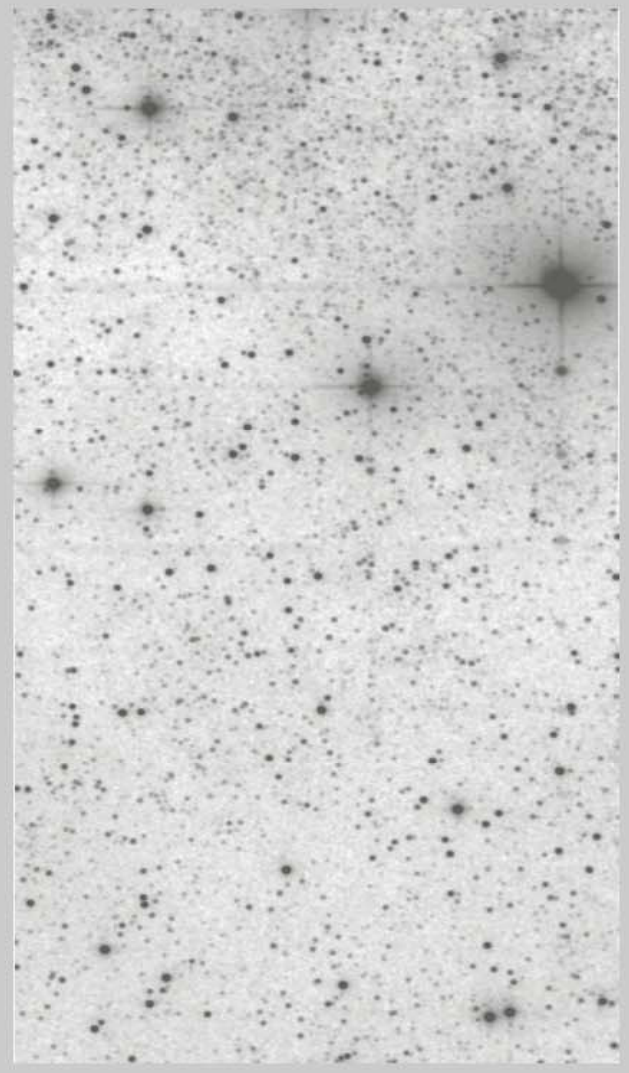
A correction is calculated for each image based on all the differences between it and its neighbors (an approximation to a least squares fit to the difference data with brightness outlier pixels excluded). The correction is currently a plane but could be a higher order surface.

This is done for all images, then half the correction determined is applied (to a parameter database; equivalent numerically to applying it to the images).

The process is iterated until step differences for all images becomes small.

Montage Background Correction Results





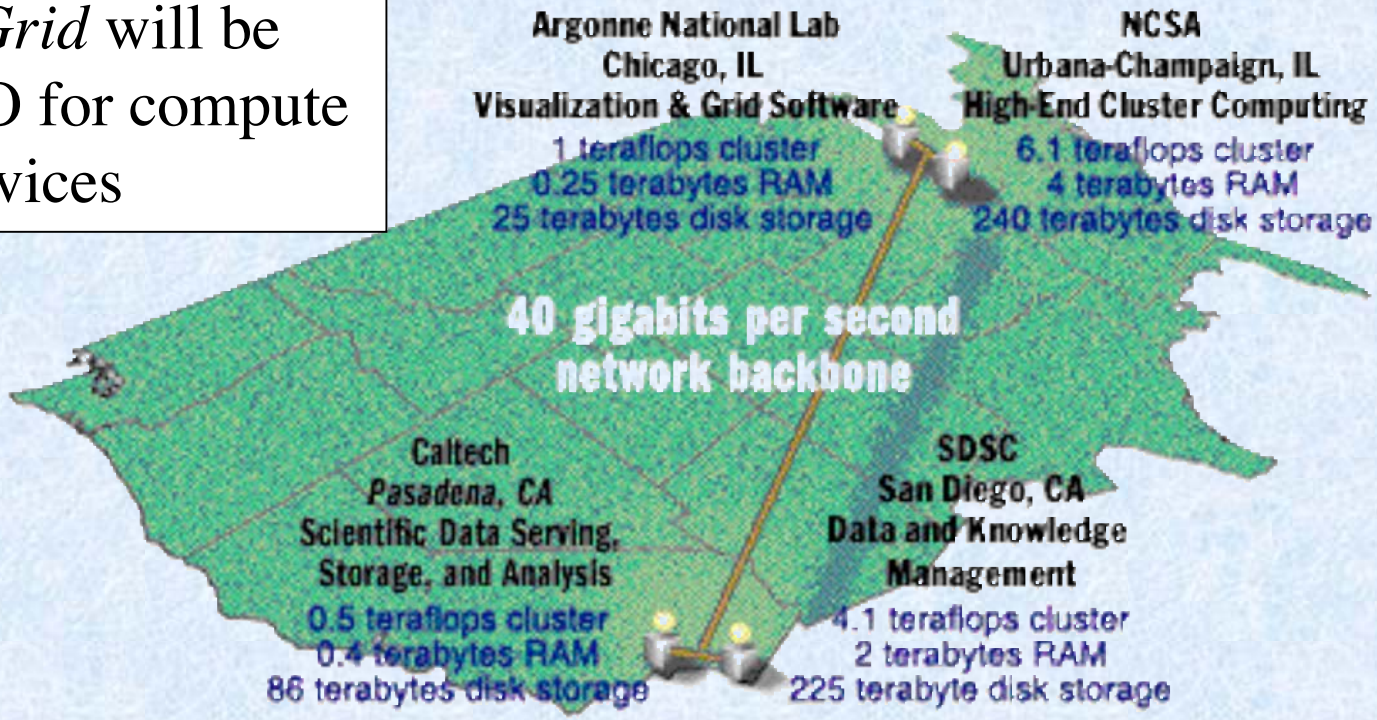
Reprojected
Background Corrected
Images

Deployment of Montage

Performance Goal:
Sustain a throughput of at least 30 square degrees per minute

Deploy operationally on the *TeraGrid* by January 2005. *TeraGrid* will be used by NVO for compute intensive services

Plus PSC (6.3 teraflops, 512 GB RAM, 150 terabytes disk storage)



Outline

- Virtual Observatories
  - Motivation
  - National Virtual Observatory
- yourSky Custom Mosaic Server
- Montage: High Science Quality Mosaics
- **Visualization**
  - **High Performance / Powerwall**
  - Web-Based / Desktop Access
- Computational Grids

Electronic Light Table (ELT)

Objective 1: To develop highly scalable interactive software for the visualization of massive, distributed astronomical datasets (both images and catalogs).

Objective 2: To illustrate to the community that the technology exists to move some traditionally catalog domain astronomical research into the image domain.

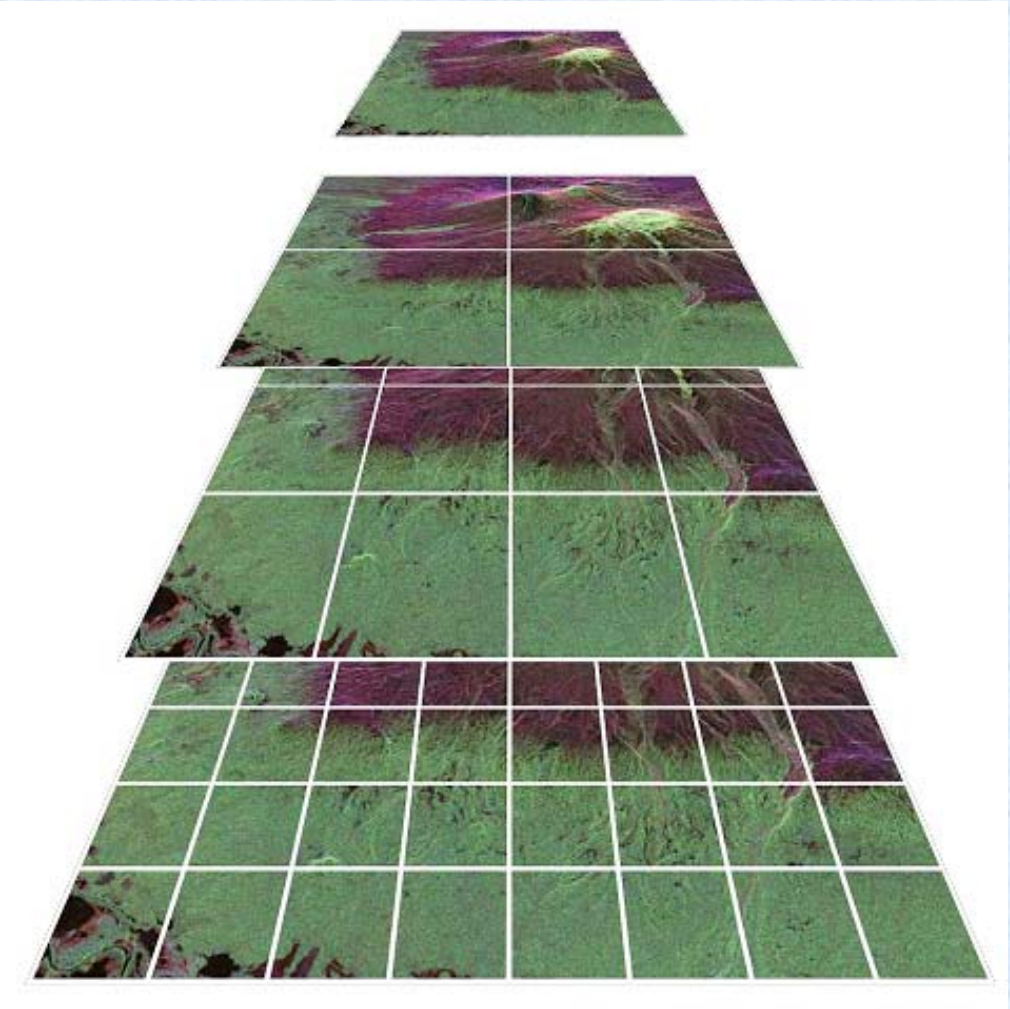
Intended for:

- Collaborative research
- Auditorium presentations
- Museums and planetariums
- Education and public outreach

Large Scale Data Input

Scalable IN:

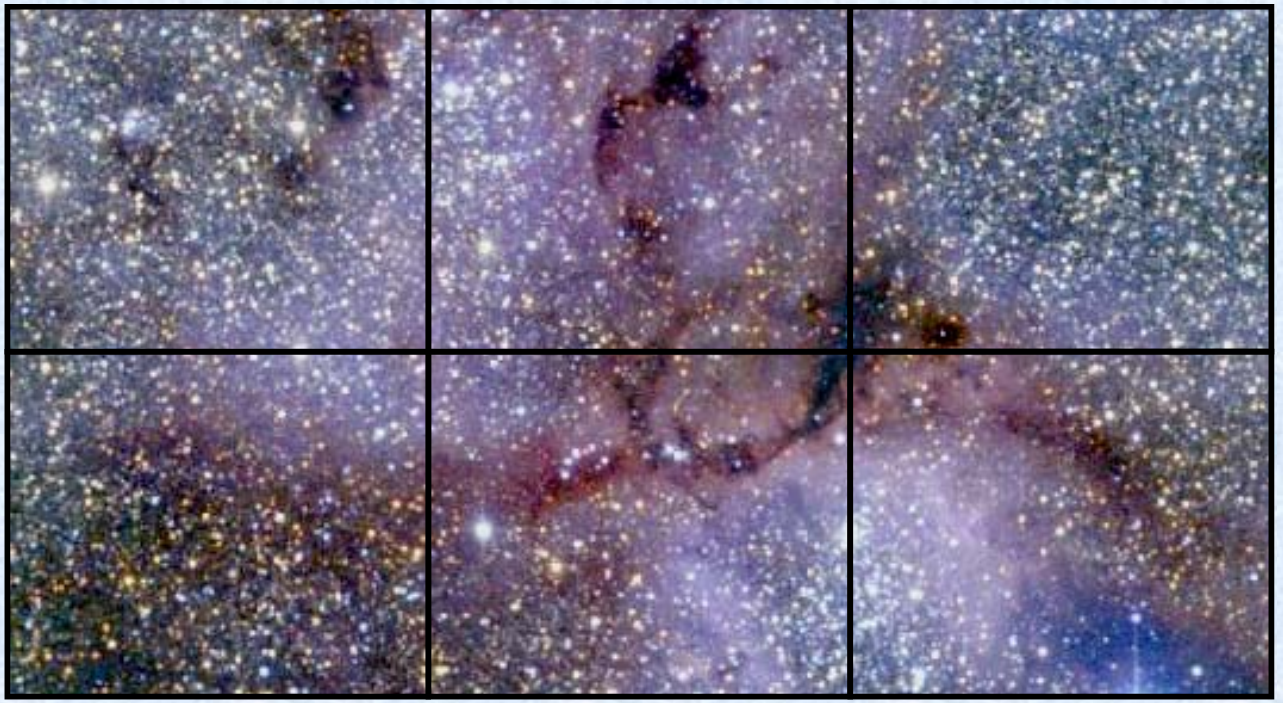
- Hierarchical, tiled data storage
- Intelligent data caching
- Rapid data access at any location in the image and at any zoom level.
- Support for most popular image formats, including the FITS image format, widely used in the Astronomy community.



Large Scale Data Output

Scalable OUT:

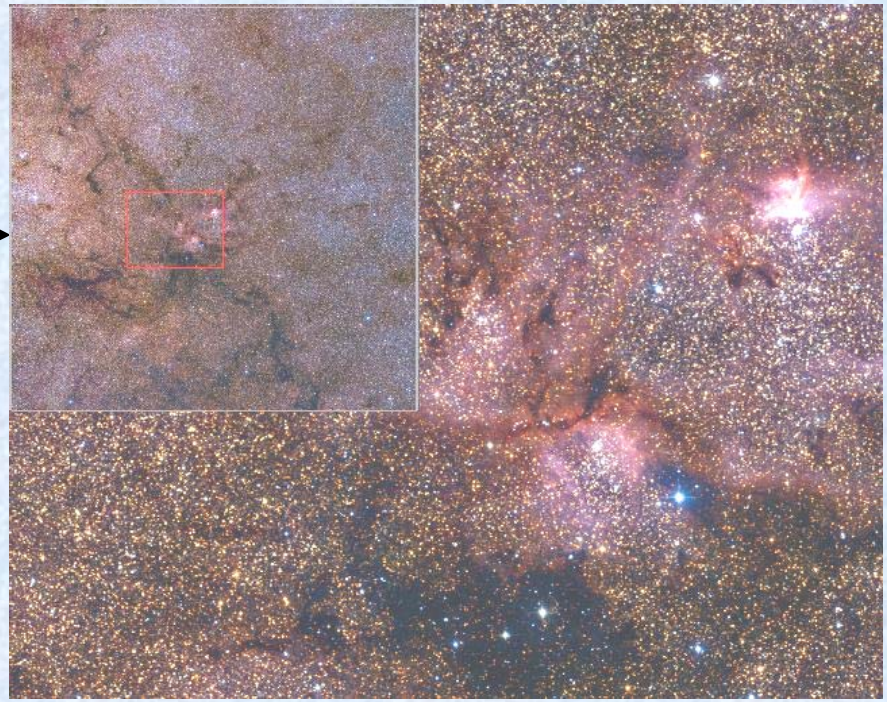
- Single-screen or multi-screen Powerwall display.
- Software synchronization of the screens with MPI.
- Example, 3x2 Powerwall display (3840 x 2048 pixels):



Large Image Navigation

- Smooth variable-speed pan/zoom with mouse or keyboard control.
- Or enter an image or sky coordinate to jump there quickly.
- Global Map View shows current location.

Global Map View →



"Data-Agile"

Pixel to Sky Coordinates

- User may jump to a specific location in the image by specifying any of the following:
  - image pixel coordinates
  - right ascension/declination
  - galactic longitude/latitude
  - ecliptic longitude/latitude
- User may also get any of these values for a pixel in the image by simply clicking on the pixel at any zoom level.

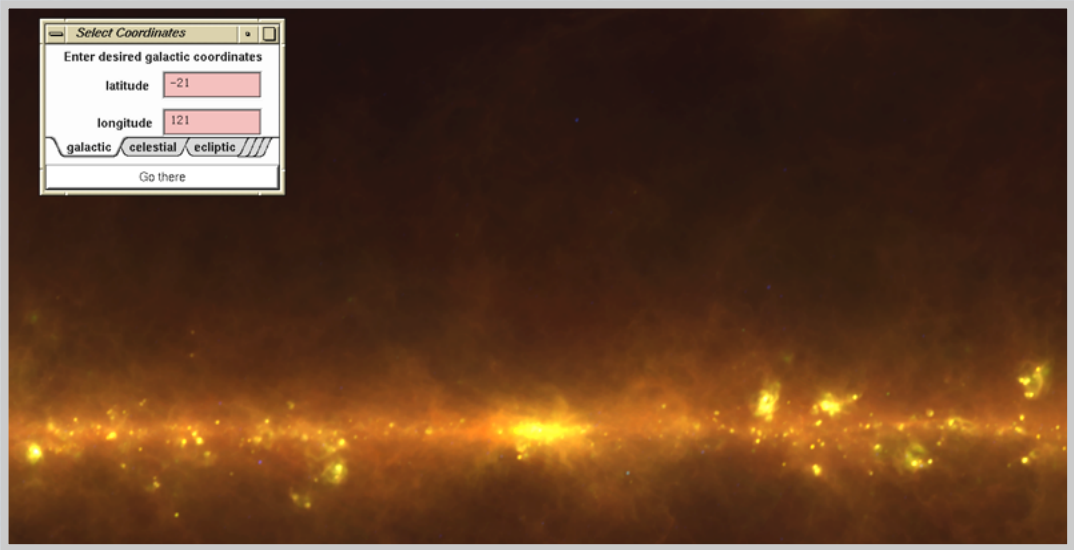
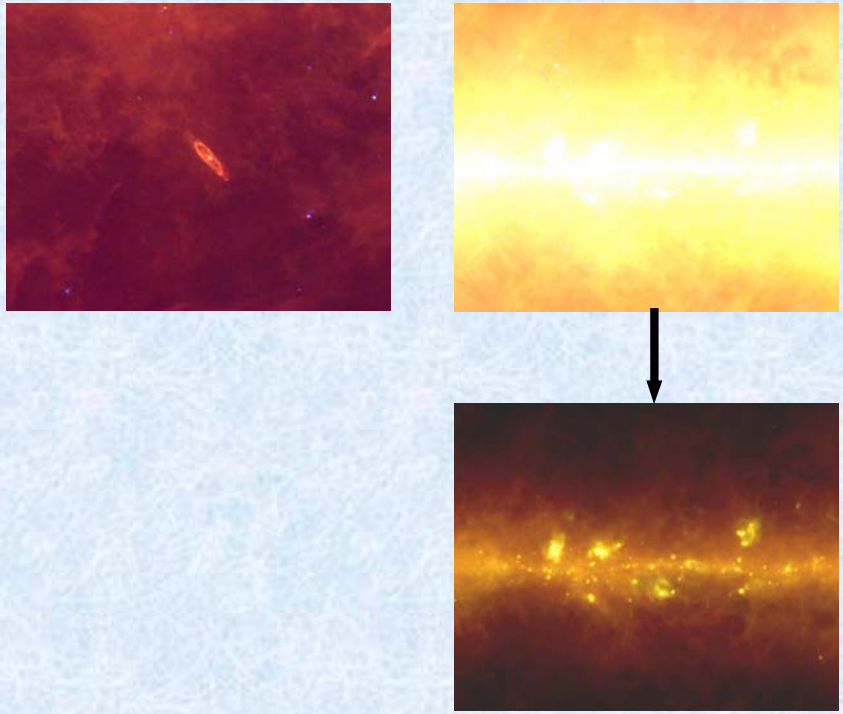


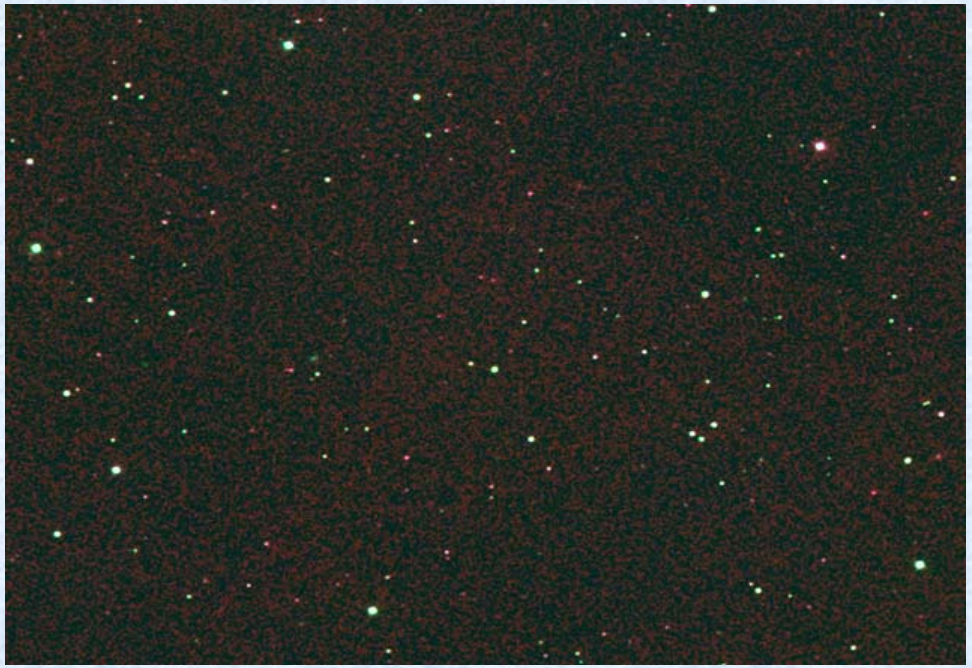
Image Enhancement

- Run-Time brightness and contrast adjustment can be applied to all three video channels (red, green and blue), or to any one channel individually.
- Brightness and contrast selectable with keyboard interface or GUI
- Example: Andromeda and the center of the Milky Way need vastly different brightness and contrast settings for optimal viewing (as illustrated below).



Multi-Spectral Viewing

- Run-Time selection of bands to be mapped to each of the R, G and B video channels.
- In screen capture below, one 2MASS band is mapped to the red gun and DPOSS is mapped to green and blue. Stars that appear red are in the infrared only, those that appear cyan are in the optical only and those that appear white are present in both.



Vector Overlays

- Support for vector overlays that pan/zoom in concert with the image.
- Overlay shape and size can be fixed (user-selectable) or determined by the values in any column in the catalogs.

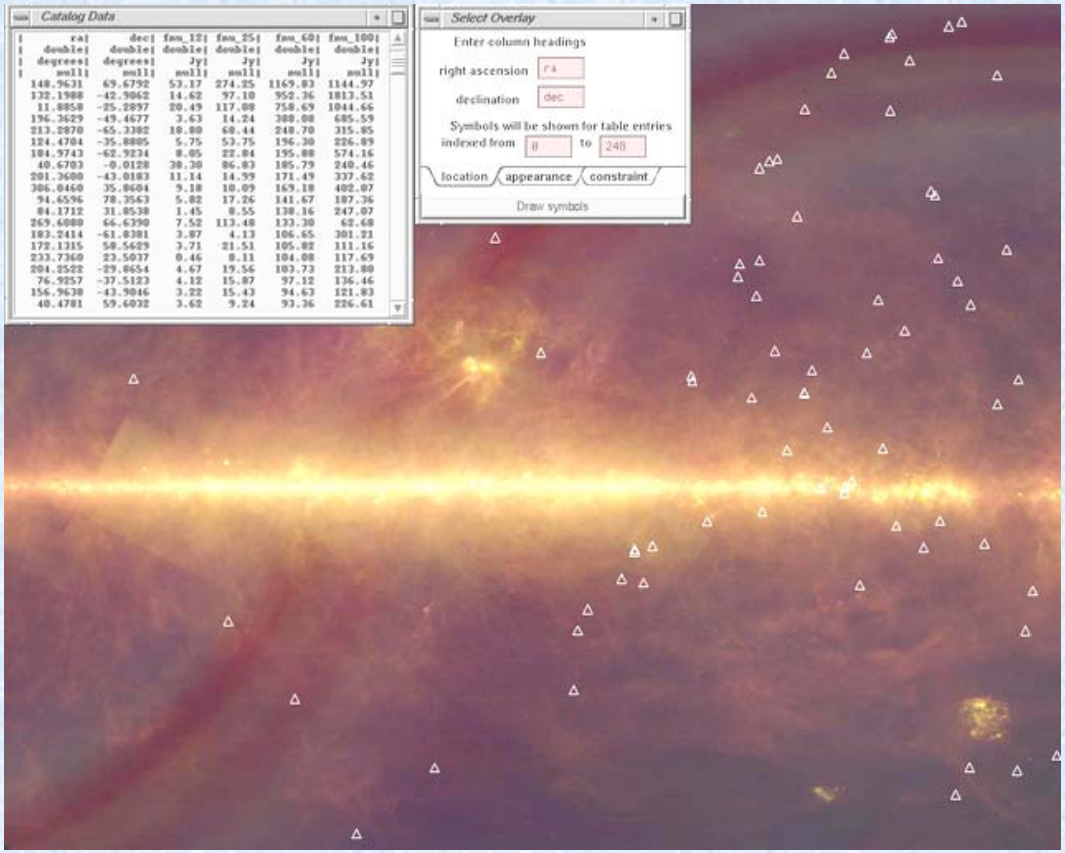
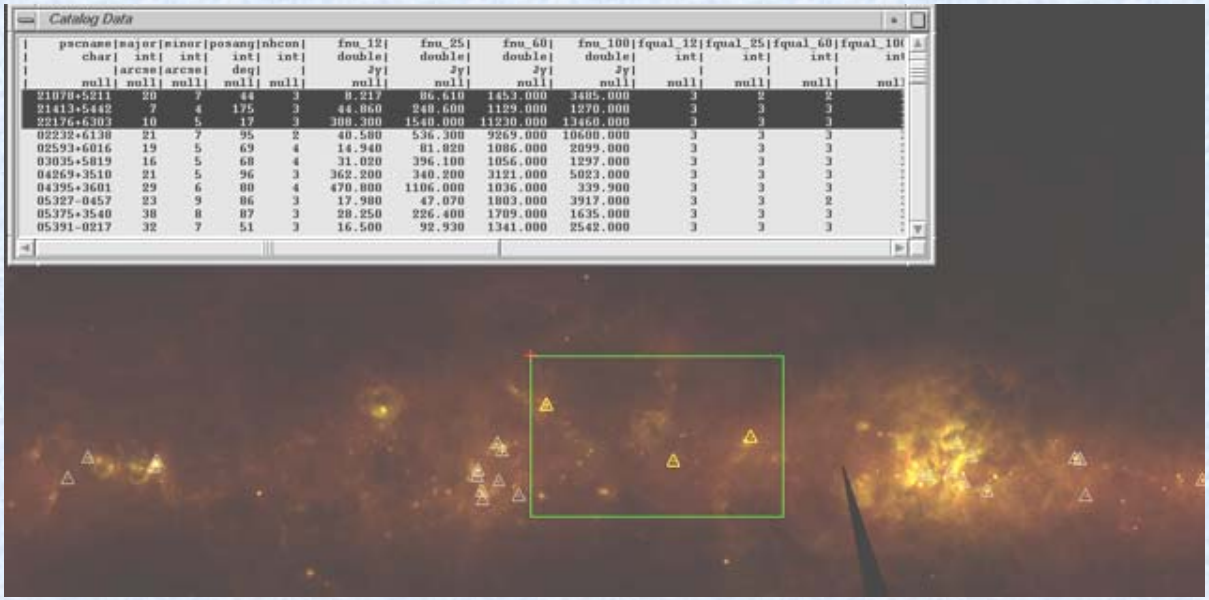


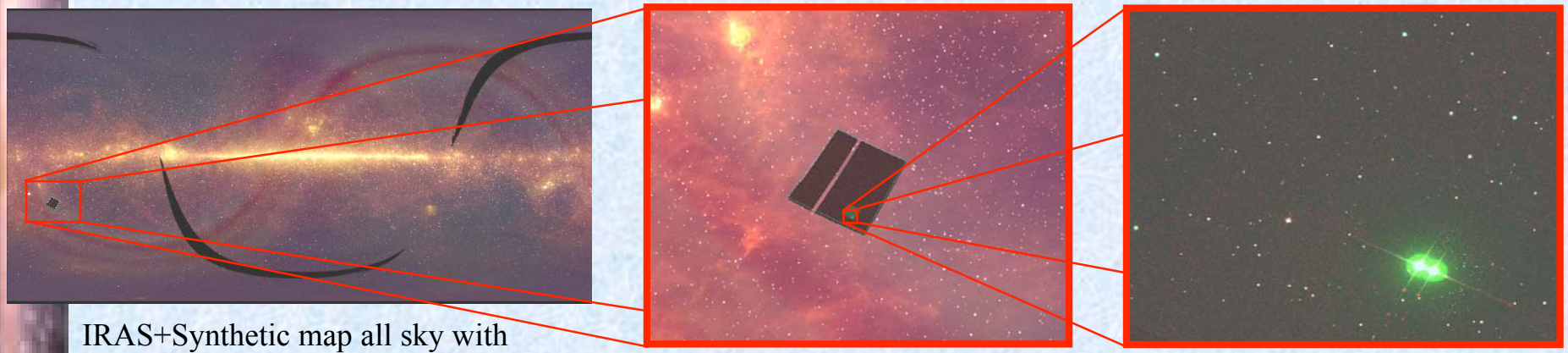
Image-to-Catalog Relation

- Relating a location in the images to catalog entries for celestial objects in the proximity and vice versa.
- **Image to Catalog:** User may select a region of the sky and see the catalog entries for those objects in that region highlighted both in the image and in the catalog window.
- **Catalog to Image:** User may select a catalog entry and see that object highlighted in the image or jump to the position of that object in the image.



Automatic Dataset Compositing

- Fully automated **run-time** compositing of multiple datasets, correctly positioned based on pixel resolution and latitude/longitude at a corner.
- Any number of datasets may be composited (although performance is degraded slightly with each one added).
- Allows user to do such things as view high-resolution insets of particular celestial objects or regions of the sky overlaid on top of lower resolution imagery of the whole sky.
- Example below shows screen captures of four composited datasets (IRAS, Hipparcos synthetic map, 2MASS, and DPOSS).



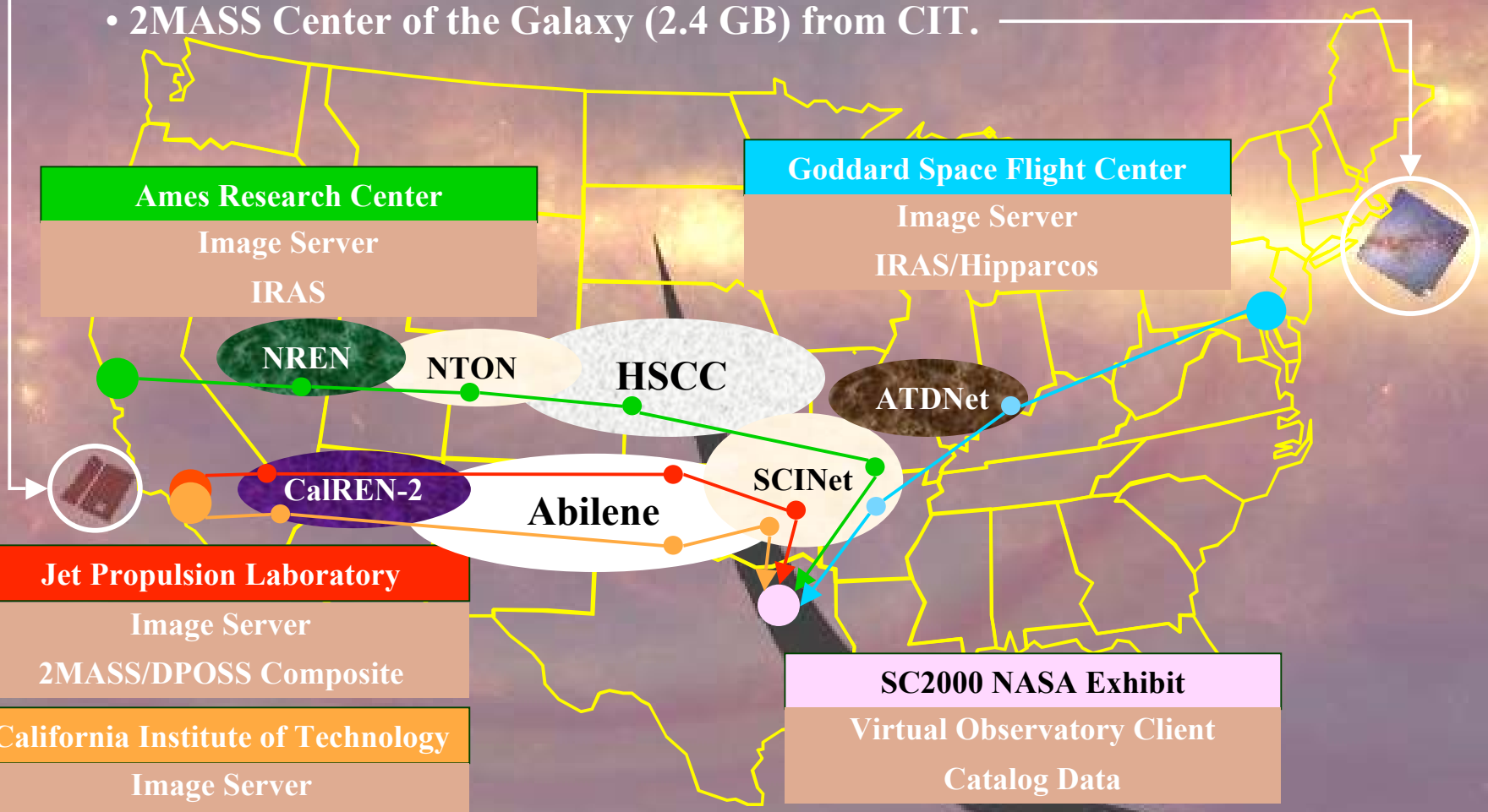
IRAS+Synthetic map all sky with 2MASS+DPOSS high resolution inset

Arc minute resolution

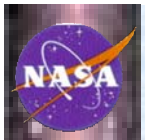
Arc second resolution

Digital Sky Virtual Observatory Demonstration at NASA Exhibit, SC2000, Dallas, TX, Nov. 4-10.

- IRAS (0.6 GB) served from GSFC and ARC.
- DPOSS/2MASS composite (10 GB) from JPL.
- 2MASS Center of the Galaxy (2.4 GB) from CIT.



Joseph Jacob, JPL, Nov. 2000



Electronic Light Table Video

Outline

- Virtual Observatories
  - Motivation
  - National Virtual Observatory
- yourSky Custom Mosaic Server
- Montage: High Science Quality Mosaics
- **Visualization**
  - High Performance / Powerwall
  - **Web-Based / Desktop Access**
- Computational Grids

Graphical Front-End to yourSky

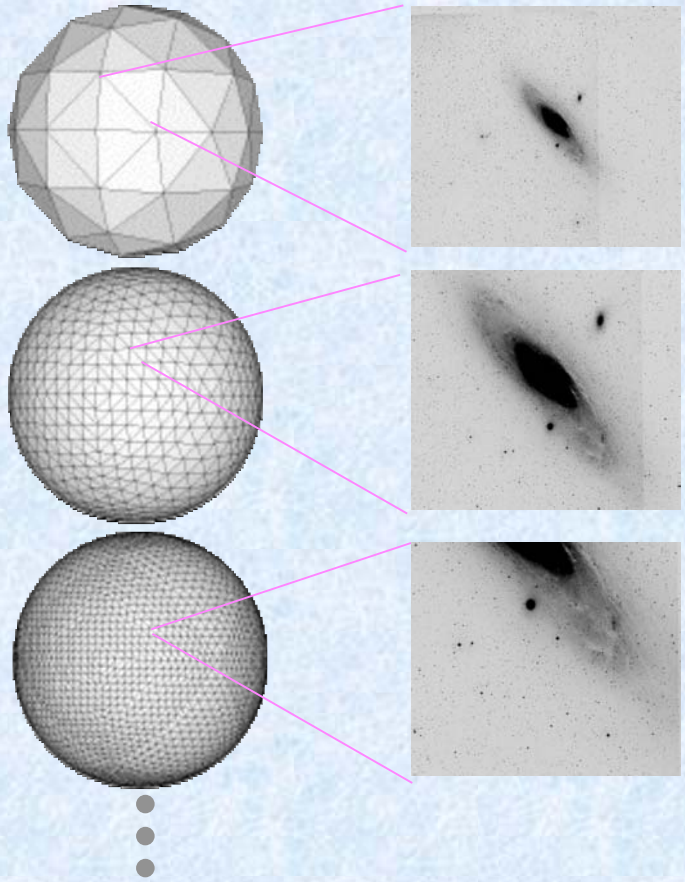
Web-Based Pan/Zoom Engine

- **All-sky browsing at medium resolution.**
- **Efficient Navigation:** Either click to re-center or zoom or enter Right Ascension (longitude), Declination (latitude) and a zoom level to jump to the desired view.
- **Multi-Spectral Viewing:** View gray scale image or map any member dataset to red, green, or blue for a color image.
- **Catalog Overlays:** Plot catalog objects overlaid on top of the image.
- **Integrated with yourSky mosaic engine:** Click a link to submit a yourSky mosaic request for the current view.

Graphical Front-End to yourSky

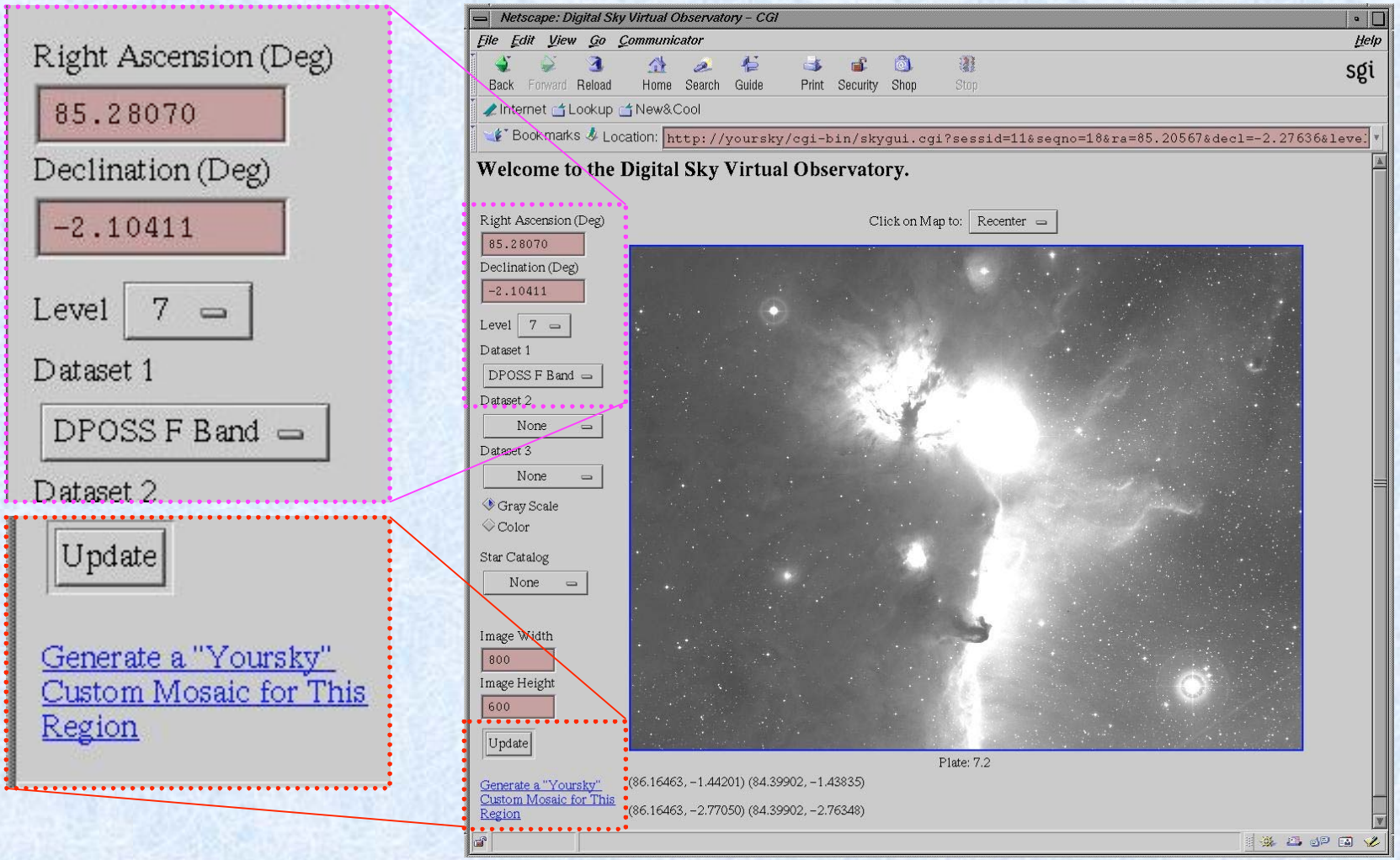
Architecture

- Hierarchical Triangular Mesh
- Single tangent plane at each vertex
- Interactive all-sky browsing
- Minimize projection distortion



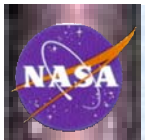
Graphical Front-End to yourSky

Sample Screen Capture



Outline

- Virtual Observatories
  - Motivation
  - National Virtual Observatory
- yourSky Custom Mosaic Server
- Montage: High Science Quality Mosaics
- Visualization
  - High Performance / Powerwall
  - Web-Based / Desktop Access
- **Computational Grids**

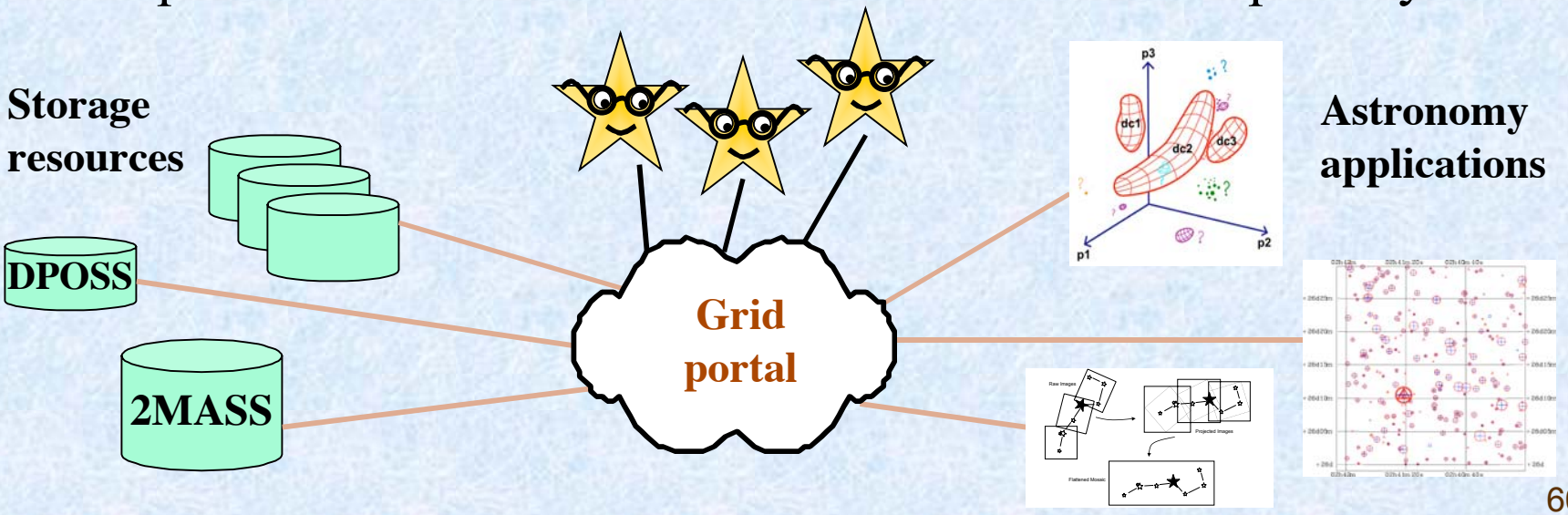


Computational Grids

- Provide the infrastructure for managing massive, distributed data collections, computational resources, and compute-intensive applications
- TeraGrid (NSF), Information Power Grid (NASA), others
- Based on Globus Toolkit and Grid Security Infrastructure
  - Execute and monitor remote applications
  - Secure file transfers: GridFTP, etc.
- OGSA (Open Grid Services Architecture) will combine Web Services and Grid technologies
- Storage Resource Broker (SRB) used to manage massive data collections and replicas

Grid and NVO: A Natural Match

- Provide NVO astronomers access to the large astronomy datasets and high power processing of the data.
- Provide the middleware glue for a nationally distributed collection of image, object archives, computational resources, and high speed networks.
- Help motivate and drive the Grid data intensive capability.

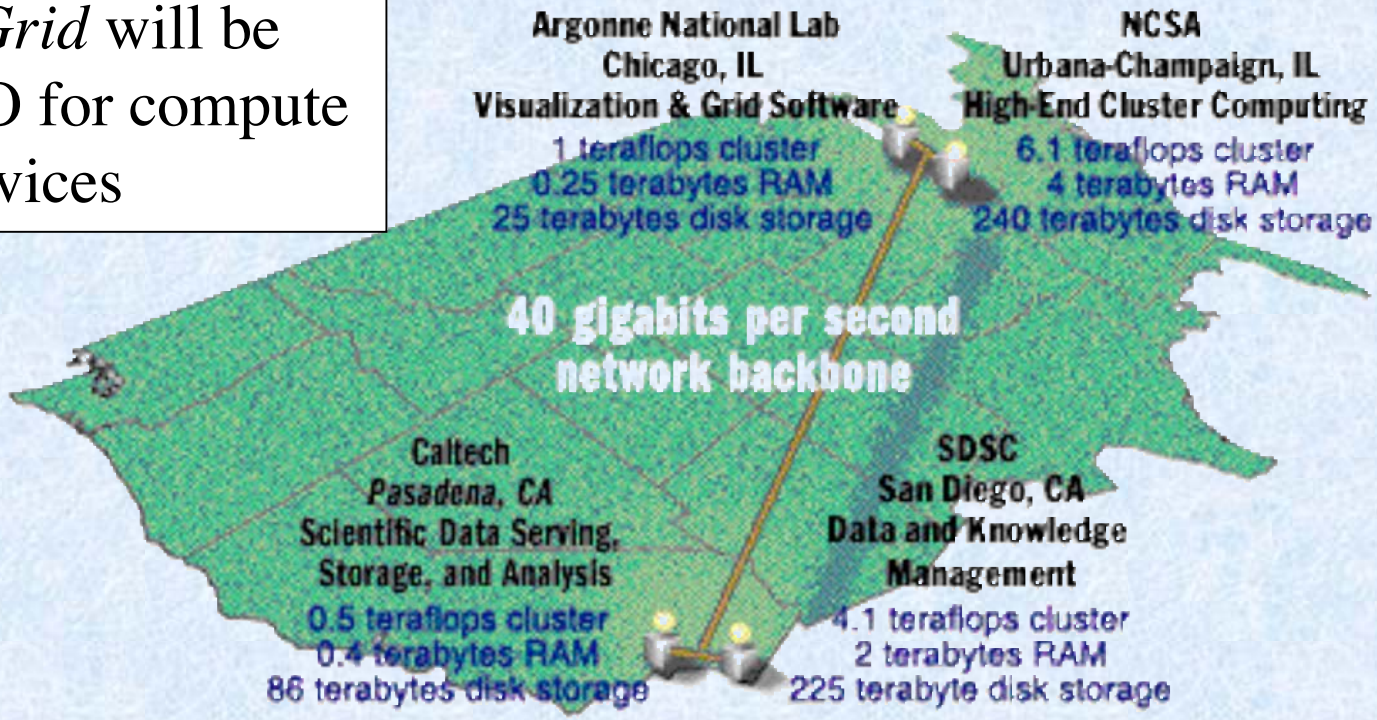


Deployment of Montage

Performance Goal:
Sustain a throughput of at least 30 square degrees per minute

Deploy operationally on the *TeraGrid* by January 2005. *TeraGrid* will be used by NVO for compute intensive services

Plus PSC (6.3 teraflops, 512 GB RAM, 150 terabytes disk storage)

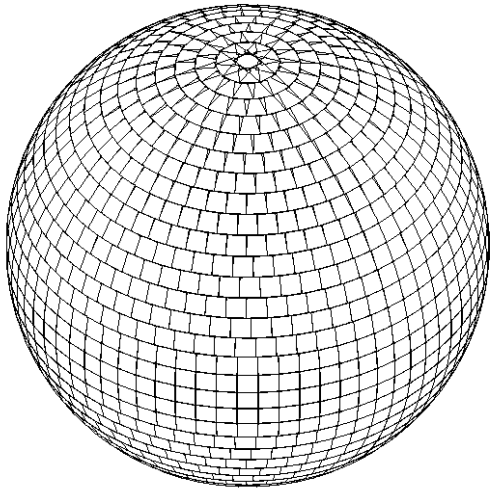


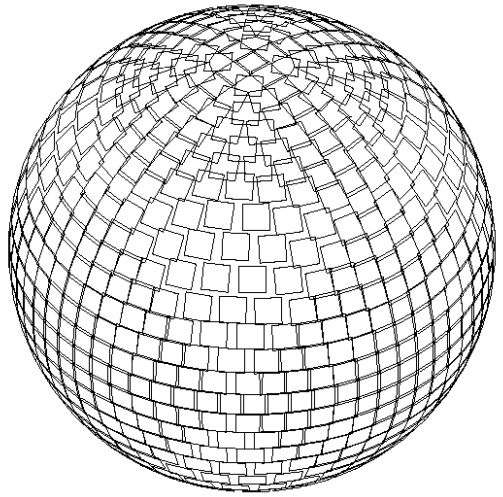
yourSkyG: Deploying yourSky on the Information Power Grid (IPG)

- IPG: NASA's computational grid infrastructure.
- yourSkyG: Globus-enabled version of yourSky
- Launch yourSky mosaicking code on the Grid instead of on local machine
- Dramatic improvements in the size and number of mosaic requests we can handle
- Construct a new set of plates accessible with the yourSky browser
  - All-sky coverage
  - Full 1 arc second resolution
  - High science quality

“Hyperatlas” Partnership

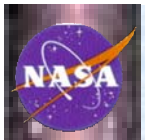
Collaboration between Caltech, SDSC, JPL, and the Astronomy domain experts for each survey.





Objectives:

- Agree on a standard layout and grid for image plates to enable multi-wavelength science.
- Share the work and share the resulting image plates.
- Involve the science community to ensure high quality plates are produced.



Summary

- The National Virtual Observatory (NVO – http://us-vo.org) will emphasize many distributed components for data archive, access, federation, processing, and visualization.
- One component of the NVO is an image mosaic service like yourSky (http://yourSky.jpl.nasa.gov) and Montage (http://montage.ipac.caltech.edu).
- The Electronic Light Table (ELT) provides high performance visualization on Powerwall displays.
- Web-based browsing is also available for desktop access.
- Grid technologies will play a key role in the NVO.

Questions?